# UNIQUELY QUANTIFYING HIGHLY SIMILAR RNA TRANSCRIPTS AT THE SINGLE MOLECULE LEVEL

A Dissertation
Presented to
The Academic Faculty

By

Gable Marsh Wadsworth

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Physics

Georgia Institute of Technology

December 2018

**UNIQUELY QUANTIFYING HIGHLY SIMILAR RNA TRANSCRIPTS AT THE SINGLE MOLECULE LEVEL**

Approved by:

Dr. Harold Kim, Advisor
School of Physics
*Georgia Institute of Technology*

Dr. James C. Gumbart
School of Physics
*Georgia Institute of Technology*

Dr. Peter Yunker
School of Physics
*Georgia Institute of Technology*

Dr. William Croft Ratcliff
School of Biological Sciences
*Georgia Institute of Technology*

Dr. Roger Wartell
School of Biological Sciences
*Georgia Institute of Technology*

Date Approved: October 24, 2018

You know, to be able to do something great in your life, you're gonna have to realize your failures. You're gonna have to embrace them and figure out how to overcome it.

*Dave Chappelle*

This dissertation is dedicated to my grandfather Dr. Gilbert Henry Koch and my uncle Dr. David Gilbert Koch who have been my inspiration for science since I was a child. I also dedicate this to my mother Kathleen Koch Almond and Bich Van McIntyre since because of their efforts I have written this dissertation.

## ACKNOWLEDGEMENTS

I would like to thank my Advisor, Harold Kim, for the support he has provided throughout my career as his student.

I would like to thank my committee for taking the time to review this dissertation.

I would like to thank my labmates in particular Rasesh Parikh, Jim Waters, Bo Broadwater, Jiyeun Jeong, and Derek Hart for the many conversations that have led to strengthening my understanding of the many concepts presented in this dissertation. I would especially like to thank Jim for the occasions he took the time to present simple coding solutions to problems that confounded me and for the amazing baked goods. I would like to thank Bo for helping me through many difficult things in and out of the lab and for being the one person I could count on to understand my sense of humor.

I would like to acknowledge the many friends I have made throughout this journey for their support that has helped me make it to this point.

I would like to thank my family for their support and understanding.

## TABLE OF CONTENTS

## LIST OF TABLES

## LIST OF FIGURES

## SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| RNA | RiboNucleic Acid |
| rRNA | ribosomal RNA |
| mRNA | messenger RNA |
| ncRNA | non-coding RNA |
| miRNA | micro RNA |
| RNase | RiboNuclease |
| DNA | DeoxyriboNucleic Acid |
| nt | nucleotides |
| cDNA | complementary DNA |
| DNase | DeoxyriboNuclease |
| LNA | Locked Nucleic Acid |
| ELF | Enzyme Labeled Fluorescence |
| FISH | Fluorescence In Situ Hybridization |
| smFISH | single molecule FISH |
| sFISH | single probe FISH |
| mpFISH | multiple probe FISH |
| FRET | Fluorescence Resonance Energy Transfer |
| HCR | Hybridization Chain Reaction |
| HILO | Highly Inclined Laminated Optical sheet |
| TIR | Total Internal Reflection |
| PCR | Polymerase Chain Reaction |
| RT-qPCR | Reverse Transcriptase - quantitative Polymerase Chain Reaction |
| YFP | Yellow Fluorescent Protein |
| RNA-seq | RNA sequencing |

| | |
|---|---|
| BFP | Back Focal Plane |
| AOTF | Acousto-Optical Tunable Filter |
| $\theta_c$ | critical angle for TIR |
| BB | BroadBand mirror |
| Di | Dichroic mirror |
| $f$ | focal length of a lens |
| $\lambda$ | wavelength of light |
| $\omega_0$ | $\frac{1}{e^2}$ width of a Gaussian beam profile |
| $MFD$ | Mode Field Diameter of a fiber |
| $T$ | coupling efficiency |
| $\omega_x, \omega_y$ | $\frac{1}{e^2}$ widths of a Gaussian beam profile for $x, y$ |
| RF | Radio Frequency |
| $\theta_B$ | Bragg angle |
| $m$ | order of diffraction |
| $n$ | index of refraction |
| $\Lambda$ | wavelength of an acoustic wave |
| $N_\lambda$ | number of resolvable wavelengths |
| $\theta_d$ | beam divergence |
| $\theta_{AOTF}$ | total scan angle of the AOTF |
| $OD$ | Optical Density |
| $f_s$ | focal length of a scan lens |
| $f_t$ | focal length of a tube lens |
| $f_o$ | focal length of the objective lens (working distance) |
| $d_1$ | distance between 2D-galvo and scan lens |
| $d_2$ | distance between scan lens and tube lens |
| $d_3$ | distance between tube lens and objective lens |
| $y$ | beam width |

| | |
|---|---|
| $y_o$ | beam width of a beam collimated out of an objective |
| EGFP | a green fluorescent protein |
| mCherry | a red fluorescent protein |
| DAQ | Data acquisition |
| LED | light emitting diode |
| UV | ultraviolet light |
| $D$ | diameter of a beam |
| $DoF$ | Depth of Field |
| $\lambda_e x$ | excitation wavelength |
| $\lambda_e m$ | emission wavelength |
| $NA$ | Numerical Apeture |
| $M$ | Magnification |
| $d_{res}$ | Abbe diffraction limit |
| $\theta_{\frac{1}{2}}$ | half apeture angle of an objective |
| $Pixel_{xy}$ | size of pixel in the sample plane |
| $Pixel_z$ | z-step size between image planes |
| CCD | Charge Coupled Device |
| EMCCD | Electron Multiplying Charge Coupled Device |
| EM | Electron Multiplying |
| $r$ | ratio of the difference in pixel intensity for each polarization to the total intensity |
| $I_p$ | Intensity of parallel polarized light |
| $I_s$ | Intensity of perpendicularly polarized light |
| $I_{total}$ | sum of $I_p$ and $I_s$ |
| PSF | point spread function of an optical system |
| N | an integer |
| $J$ | Bessel function of the first kind |

| | |
|---|---|
| $k$ | wavenumber |
| $r_a$ | radius of the aperture |
| $\theta_o$ | angle of observation |
| FWHM | full width at half maximum |
| $I_0$ | maximum intensity |
| ROI | region of interest |
| GUI | graphical user interface |
| $\sigma$ | the standard deviation of a Gaussian profile |
| LoG | Laplacian of a Gaussian |
| $O$ | observed image |
| $I$ | true image |
| $\nu_i$ | weight coefficient |
| $v(x)$ | gain term |
| $z(x)$ | noise term |
| $G_x$ | a kernal for a Sobel filter |

## SUMMARY

Understanding RNA expression is critical to understand the connection between phenotype and genotype in living cells. Popular methods to characterize the distribution of RNA transcripts are all currently limited by the length of the transcript of interest. In the case of PCR based methods, uniquely identifying highly similar transcripts requires particular effort in amplification, depletion or read depth to avoid bias and most assays cannot detect short transcripts. In the case of In Situ Hybridization based detection, the limitation is based on the length of the unique features of the sequence. Short sequences are unable to be detected because they do not have adequate length to support multiple probes to obtain a high signal to noise ratio. The work presented in this thesis introduces new methods to quantify RNA transcripts that address these fundamental concerns in particular (i) how well can a short single probe quantify RNA transcripts and (ii) how to uniquely identify RNA transcripts that are highly homologous.

RNA Fluorescence In Situ Hybridization (FISH) is a popular technique used to quantify the number of RNA transcripts inside single cells and has been critical in understanding the two-state model of gene regulation. In this method, an RNA transcript is targeted with multiple 20-50nt DNA oligo probes so that there are 20-50 probes bound to each transcript requiring that a sequence be long ($\geq$200nt). The requirement of multiple labeling is due to the need for increased intensity of detected spots and the desire to obtain the absolute number of transcripts. However, in many cases it is adequate to use relative quantification of the transcript level to reveal a difference between conditions or to demonstrate heterogeneity between single cells. In these cases, it is possible to use a short singly labeled DNA oligo probe to quantify the relative number of transcripts in a population of cells. This method is demonstrated in Saccharomyces Cerevisiae (yeast) on the exogenous gene yEVenus (YFP) and the well characterized endogenous gene KAP104. This method also reveals the subcellular localization of the non-coding transcript TLC1. It is demonstrated that this method

can be used to enable Fluorescence Resonance Energy Transfer (FRET) detection of RNA by designing a pair of probes to target neighboring binding sites. This provides unambiguous identification of transcripts, can report the colocalization of two sequences, and can characterize the detection efficiency of a single probe.

While a single probe can discriminate between the identity of transcripts that are homologous except for a short unshared sequence, it is unable to discriminate between sequences such as splicing variations that are completely homologous to the canonically spliced isoform. Combining RNA FISH with Hybridization Chain Reaction (HCR), it is observed that a single mismatch at the beginning of the stem of the hairpin loop prevents HCR from progressing in the case of probes designed to target yEVenus. It is also demonstrated that these probes are metastable in a negative control. This has been used to distinguish intron retention and circularized isoforms of SUS1 in yeast at elevated growth temperature ($37\,^{\circ}\mathrm{C}$). A further extension of this method incorporates a universal probe that is designed to bind a pair of unlabeled HCR probes. These probes also incorporate a short additional toehold so that a third strand displacement reaction can deconstruct the amplified DNA polymer and enable sequential labeling of exon-exon sequence junctions in the NURF1 gene in *Caenorhabditis Elegans*. This sequential barcoding method is used to understand the spatial distribution of NURF1 isoforms in the *C. Elegans* germline.

# CHAPTER 1
# INTRODUCTION

The need to understand the connection between the genotype of an organism and its phenotype is the driving factor behind the need to quantify RNA transcripts in cells. In a broad context, the transcriptome represents the direct connection between the genome and the proteome since each protein is the result of translation of an RNA transcript. However, the transcriptome is not limited to cases that follow the Central Dogma of Molecular Biology since many transcripts also serve regulatory functions, which do not require protein expression as the terminal goal of a transcript. RNA transcripts have been studied in a number of ways beginning with ensemble approaches such as the northern blot[1] and quantitative reverse transcription polymerase chain reaction (RT-qPCR)[2, 3]. In both of these cases, a population of cells is lysed and the total amount of RNA transcripts in all cells is isolated. Transcripts are then identified using either microarrays[4] or RNA-seq, which has become the standard for sequencing based RNA detection[5]. There are problems in using RNA-seq to detect transcripts, which include sequencing bias leading to difficulty comparing transcript levels and technical artifacts that make it difficult to separate real differences in sequences from systematic error[6, 7]. Contamination such as changes in salt concentration or pipetting error can cause variability in these methods[8]. Commercially available systems for qPCR suffer because they are unreliable for short sequences below 75 nucleotides. The ability of PCR based methods to detect uncommon transcripts is limited because the amount of transcripts needed for reliable amplification is high. This makes it nearly mandatory to deplete the RNA sample of rRNA transcripts, which can also introduce error or bias if the sample preparation is not carefully performed. Finally, purification of RNA to perform RNA-seq can copurify inhibitors providing further difficulty to this technique[9]. Despite the difficulty in performing these techniques on ensembles of cells,

RNA-seq has also been demonstrated for single cells[10, 11, 12, 13, 14, 15]. Despite being able to provide information about the entire transcriptome of a cell, the limitations of PCR methods make it difficult to infer anything about transcripts which are low in copy number at around 0-15 copies. While the advancements made in these techniques allow single cell quantification, they do not retain spatial information in most cases and the cost can be inhibitory. There should be a balance between cost, throughput and spatial resolution.

To perform RNA-seq, the total RNA in a population of cells[16] or a single cell[17] is converted into a cDNA library by destroying the sample and purifying the RNA. This RNA library is then sequenced via a next generation sequencing method, sequencing by synthesis (Illumina) or sequencing by ligation (SOLiD). The depth of read of the data set determines whether or not it is possible to detect short or rare transcripts since differential gene expression requires much less reads to detect. This difficulty is compounded when the organism is less studied because expression levels might not be known. Furthermore, it is necessary to develop Bioinformatics tools that can detect things like isoforms or novel splicing such as circular RNA, since these might be discarded as artifacts in data processing. RNA-seq is becoming the gold standard for gene expression, however, many of the techniques using RNA-seq also still depend on other methods to obtain the spatial distribution of transcripts and for validation of transcript levels.

The most common technique that provides spatial information and is both lower in cost and throughput is single molecule Fluorescence In Situ Hybridization(smFISH)[18]. This technique involves targeting an RNA with many complementary DNA oligo probes labeled with one or multiple fluorophores. Typically it requires 20-30 distinct probes to obtain a good signal to noise ratio. RNA are then detected by fluorescence microscopy. Unlike qPCR methods, smFISH based techniques are non-destructive since they require fixation of the cells or tissue. This preserves the sample and allows for storage from several weeks to months. Another advantage to FISH over the PCR based methods is that there are no amplification processes necessary to develop a signal. Following fixation, the cell

is permeabilized to the probes and RNA degradation is halted with vanadyl ribonucleoside complex. This means that the structure of the cellular interior is preserved and available for interrogation by any method necessary. It is therefore possible to combine other methods with FISH, such as, immunofluorescence of a related protein or even subsequent RNA-seq. These techniques have been refined to analyze single nucleotide polymorphisms[19], obtain super resolution images of subcellular structure[20, 21] and chromosomes[22], as well as the detection of miRNA with an indirect detection via enzyme-labeled fluorescence signal(ELF) amplification combined with LNA-FISH[23]. smFISH allows for the absolutue number of transcripts to be determined for a single gene.

Detection of the transcripts of multiple genes can be accomplished simultaneously using spectrally discrete fluorophores for only up to 5 unique transcripts. To accomplish detection of more transcripts requires techniques that involve multiple sequential hybridization steps and typically include super-resolution techniques. Spectral barcoding[24] can increase the number of simultaneously detectable unique transcripts combinatorially, $\binom{n}{r}$, limited by the number of spectrally discrete fluorophores available. Spatial barcoding[25] relies on super resolution methods to determine the order of the fluorophores and the number of unique transcripts detectable scales as $n!$. In either case smFISH is rather limited in throughput compared to the exhaustive detection of RNA-seq. The closest that a FISH based detection scheme can come is temporal barcoding[26], which can detect thousands of unique transcripts through sequential steps of DNase and hybridization where each round increases the amount of unique targets detected. In the case of the methods requiring super-resolution techniques, the methods are generally time consuming and destructive to the sample and require high labeling density to obtain high quality spatial information.

smFISH can also reveal information regarding the kinetics of stochastic gene transcription, which in simple cases can be modeled by a Poisson distribution. However, the telegraph or two-state model of transcription was developed by counting nascent sites of transcription[27, 28]. These nascent sites revealed that gene expression under regulation

was stochastic and demonstrated bursts of transcription spaced by long periods of inactivity. When detected via smFISH, the large number of sub-voxel localized transcripts is observed as a significantly brighter spot and not distinguished by unique features such as introns. One of the areas that smFISH has largely been unable to address is the unique identification of transcripts that are highly homologuous. For smFISH protocols, the need for multiple probes means that unique features shorter than $\sim$20 nucleotides will not be observable. This means any variation in transcript sequence either by the splicing or transcription process is undetectable. RNA isoforms are an important part of understanding the transcriptome and certain types of isoforms can be upregulated or downregulated in cancer making it desirable to be able to extend a popular technique such as smFISH to this class of transcript. When RNA isoforms are part of a regulated process and not the failure of the function of the spliceosome, the number and identity of RNA isoforms and their pre-spliced precursor RNA can generate a velocity of transcription[29]. In order to predict the future state of the transcriptome of a cell, the time evolution of spliced and unspliced molecules can be estimated by a pair of coupled rate equations. The change in expression over time can then be represented as a velocity, which represents the transcriptional dynamics of asingle cell.

This thesis represents an effort to characterize the performance of a singly labeled DNA oligo probe for RNA detection and subsequently to leverage nucleic acid strand displacement kinetics to uniquely identify RNA isoforms in *Saccharomyces Cerevisiae* (yeast) and *Caenorhabditis Elegans*. In particular for yeast, we use a constitutively expressed exogenous gene for the yellow fluorescent protein, yEVenus, to characterize single probe performance and establish the use of Fluorescence Resonance Energy Transfer (FRET) *in situ*. We demonstrate the impact of fixative on probe detection efficiency and the utility of inclined illumination to increase signal to noise. This method required extensive testing of different optical systems, which are discussed in Chapter 2, as well as, the development of a FISH analysis pipeline, which will be discussed in Chapter 3. The method of single

fluorophore FISH will be discussed in Chapter 4. The remainder of this thesis will present unpublished efforts with two collaborations with the first being a study of the non-coding RNA TLC1 presented in Chapter 5. This transcript represents the RNA base for proteins to bind and form the holoenzyme telomerase. Our efforts reveal that TLC1 changes localization and diffusivity based on temperature sensitive mutant alleles of the pop1 and pop6 genes. Finally in Chapter 6, strand displacment based specificity is explored using yEVenus and SUS1 in yeast as well as NURF1 in *C. Elegans*. The results using Hybridization Chain Reaction (HCR) represent the first use of strand displacement as a means of obtaining single nucleotide specificity. Further, the experiments presented for circular RNA represent the first direct observation of topologically distinct RNA. This work is then combined with an additional strand displacement reaction to demonstrate a method to characterize all isoforms of an RNA using sequential hybridizations. This method requires a footprint of 20 nucleotides which far surpasses the previous limitations for detection, a minimum of 200 nucleotides.

# CHAPTER 2
# MICROSCOPY

In this section, the various configurations of the microscope will be discussed. All customization is done around a commercial inverted microscope frame (Olympus, IX-82). This frame includes a lamp, condenser, dic prism, filter wheel, and objective turret that are controlled via Micromanager[30]. Also, the frame includes a manually adjusted mirror to select between two optical paths.

## 2.1 Excitation path

### 2.1.1 Gas lasers and AOTF

The first iteration of the illumination pathway included two gas lasers, an acousto-optic tunable filter, and a fiber optic to combine multiple wavelengths into a single fiber source. The output of the fiber is collimated using a $5\,\mathrm{cm}$ focal length lens (Figure 2.1. The collimated beam is then directed into the microscope by a pair of relay mirrors and is focused on the back focal plane using a z-translation stage and a tube lens of focal length $25\,\mathrm{cm}$.

This system can be set to either achieve epi-illumination if the fiber plate in the XY translation is centered or to TIR/HILO if the position of the fiber tip is radially displaced. Total internal reflection (TIR) is achieved when the beam is displaced to the edge of the back focal plane and the beam subsequently is incident on the glass coverslip at the critical angle $\theta_C$. In TIR, the sample is illuminated by the evanescent wave, which attenuates rapidly in power and is capable of illuminating only the $100\,\mathrm{nm}$ layer near the surface. Highly Inclined Laminated Optical sheet (HILO) is a sub-critical alignment of the beam to create a thin intense inclined beam capable of z-sectioning the sample.

While versatile, this system included several points of power loss that made it difficult

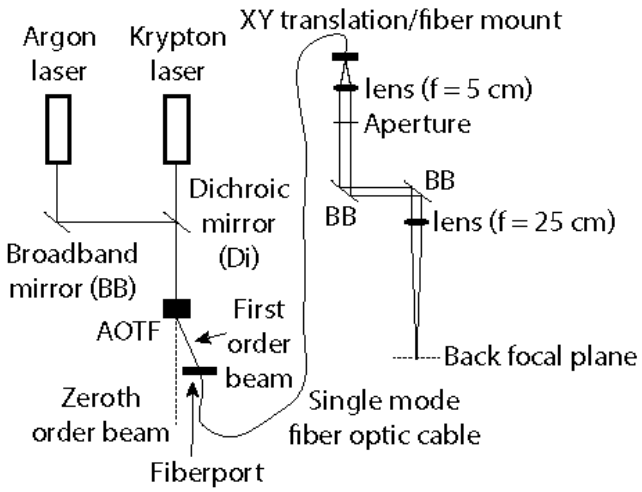


Figure 2.1: Schematic of gas laser setup. An argon/krypton laser setup is used to have six channels of excitation wavelengths; $455\,\mathrm{nm}$, $476\,\mathrm{nm}$, $488\,\mathrm{nm}$, $514\,\mathrm{nm}$, $568\,\mathrm{nm}$, and $647\,\mathrm{nm}$. These channels are selected via an AOTF and coupled to a fiber optic cable.

to acquire single fluorophore data inside of cells due to only having $0.5\,\mathrm{mW}$ of power at the back focal plane of the objective for excitation of Cy3 and Cy5.

*Coupling to a fiber in free space*

The laser was coupled to a fiber using a fiberport (Thorlabs, PAF-X-11-PC-A). This port includes XY translation via adjustment of two set screws on the top and bottom and Z translation and angular adjustment via three screws that mount the lens to the housing. The choice of lens was made using the formula relating the mode field diameter (MFD) to the focal length of the lens $f$ as well as the $\frac{1}{e^2}$ width $\omega_0$ and wavelength $\lambda$ of the gaussian beam[31].

$$MFD = \frac{4\lambda f}{\pi\omega_0} \tag{2.1}$$

The easiest way to align z-position of the lens in the fiberport is to exploit time reversal symmetry of the laser and collimate a beam exiting the lens. Subsequently, the angle of the plate should be adjusted by fixing two of the screws and making small adjustments on the third until a local maximum in output intensity is achieved.

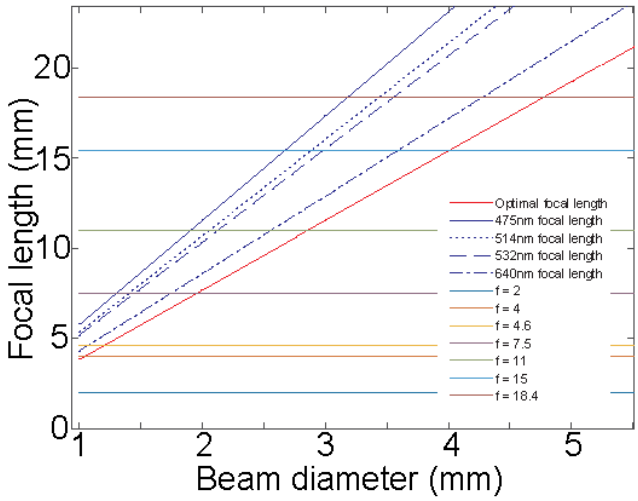


Figure 2.2: Determining the focal length for the fiberport. This plot shows the appropriate focal length of lens to couple to a laser in free space to a fiber. The red line shows the optimally matched Gaussian waist to mode field diameter (MFD) of the fiber based on the half angle of acceptance of the fiber. The blue lines show the focal length necessary to match each wavelength to the fiber. Any beam below the red line would under fill the fiber. Lines above the red line have waists larger than the MFD. At fixed beam diameter, coupling efficiency should be adequate if the focal length chosen is above the optimal focal length for each wavelength.

The theoretical coupling efficiency of this beam neglecting mechanical alignment errors can be found using the following equation[31, 32] where T is defined as the efficiency, $\omega_x$ and $\omega_y$ are the widths of the incoming beam waist, and $MFD$.

$$T = \frac{4\omega_x\omega_y MFD^2}{(MFD^2 + \omega_x^2)(MFD^2 + \omega_y^2)} \tag{2.2}$$

This represents the coupling efficiency in an ideal case and is quite difficult to achieve without very precise alignment. For the gas lasers, coupling efficiency was at best 10%, which is significantly less than the theoretical maximum.

*Aligning an AOTF*

In an AOTF, a piezo-transducer is driven by an RF signal and produces an longitudinal acoustic wave through a tellurium dioxide crystal. The regions of compression and rarefac-

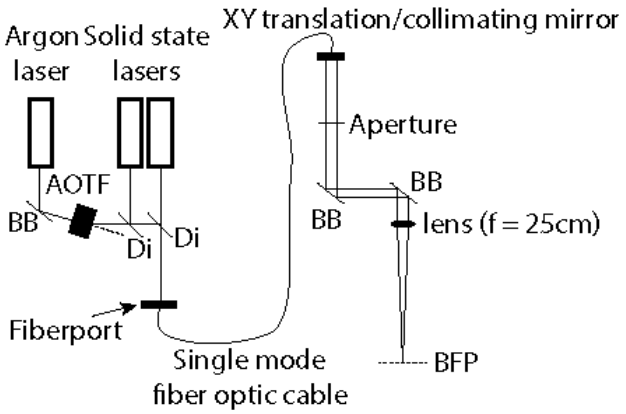


Figure 2.3: Schematic of solid state laser setup. Two solid state lasers are added to the argon laser; $455\,\mathrm{nm}$, $476\,\mathrm{nm}$, $488\,\mathrm{nm}$, $514\,\mathrm{nm}$, $532\,\mathrm{nm}$, and $640\,\mathrm{nm}$. The argon laser channels are selected via an AOTF. The solid state lasers are controlled via serial directly for the Coherent laser and through a mechanical shutter for the Oxiuss laser and coupled to a fiber optic cable.

tion serve conceptually as the slits in a diffraction grating based on the modulation of the refractive index via the photo-elastic effect. The power and frequency of the acoustic wave regulates the wavelength and intensity of light that can pass through the crystal. Typically, these types of filters can have up to 85% efficiency and switch at greater than $1\,\mathrm{MHz}$.

There are two regimes for an AOTF, if the length the laser illumination travels through the acoustic wave is thin, then this is known as the Raman-Nath interaction regime and several diffraction orders are observed. When the crystal is adequately thick, then the interaction is in the Bragg regime and only the first order beam is produced as the rest are annihilated by destructive interference.

The Bragg angle $\theta_B$ can be determined using the order of diffraction $m$, the refractive index $n$, the wavelength of the incident light $\lambda$ and the wavelength of the acoustic wave $\Lambda$[33].

$$sin\theta = \frac{m\lambda}{2n\Lambda} \tag{2.3}$$

Empirically, when aligning the AOTF, the efficiency of deflection of the laser is high enough that at $\theta_B$ the majority of the laser intensity should be shifted from the $m = 0$

fringe into the $m = \pm 1$ fringes when the AOTF is driven at the appropriate frequency and amplitude, which can be found in the manufacturer's specifications.

The resolution of an AOTF is on the order of several nanometers and this resolution also can be found in the manufacturer's specifications. The number of resolvable wavelengths $N_\lambda$ is determined by the divergence of the incident beam $\theta_d$ and the total scan angle of the AOTF $\theta_{AOTF}$.

$$N_\lambda = \frac{\theta_{AOTF}}{\theta_d} \tag{2.4}$$

One downside to using the AOTF and coupling to a fiber is the inherent cylindrical lensing due to the fact that the AOTF only affects one axis of the beam profile. This contributes to approximately an additional 50% loss in laser intensity when coupled to the single mode fiber in the setup described above.

### 2.1.2 Solid state lasers

The illumination path was modified to remove gas lasers and add solid state lasers for several reasons. Firstly, the gas laser requires routine servicing including periodically re-aligning the mirrors in the cavity and the relatively frequent need for replacement of the gas tube. Secondly, the cost of replacement of these components is similar to the cost of a solid state laser. Finally, the primary wavelengths of use had low power output compared to the solid state options.

Initially, only the laser that served to excite Cy5 was replaced (Coherent, 1185055). However, an additional laser was subsequently added for Cy3 excitation (Oxxius, LCX-532L-100). The Coherent laser is capable of accepting serial commands over USB and is controlled via Micromanager. The $532\,\mathrm{nm}$ laser did have a USB connection, however, the laser intensity is fixed at $100\,\mathrm{mW}$, which required the addition of a mechanical shutter (Thorlabs, SH10), controller (Thorlabs, SC10), and a neutral density filter with a gradient between 0.04-4.0 optical density(OD) (Thorlabs, NDL-25C-4). This filter is positioned via an xy translation mount (Thorlabs, XYFM1) with $50\,\mathrm{mm}$ of travel.

Using the AOTF, the amount of power observed at the BFP was $15\,\mathrm{mW}$ for these lasers set to $100\,\mathrm{mW}$. When removing this from the optical path, much greater coupling efficiency was achieved. The best coupling to the fiber was $55\,\mathrm{mW}$ for the $640\,\mathrm{nm}$ laser, although typically it was coupled at approximately $48\,\mathrm{mW}$ to also optimize the coupling efficiency of the $532\,\mathrm{nm}$ laser, which is coupled at $35\,\mathrm{mW}$ when optimizing for both lasers.

A final solid state laser was incorporated into the system. This laser is a $50\,\mathrm{mW}$ $473\,\mathrm{nm}$ laser (Optotronics, VA-I-xx-473), which arrived with TEM01 mode with two lobes rather than TEM00 which is a single gaussian mode. These are the Hermite-Gaussian modes. This laser produces a total power output of $25\,\mathrm{mW}$ and in general under performs the manufacturer's specifications. When attempting to couple this laser to the fiber, a maximum of $2\,\mathrm{mW}$ was achievable without significantly perturbing the coupling of the other two lasers. This laser motivated the shift from the setup using an AOTF and single mode fiber to launching the beams directly into the microscope. In that case, the power loss at the BFP was nominal at between 10-20% for all the lasers.

The $473\,\mathrm{nm}$ laser is also capable of control via TTL input and is controlled in Micro-manager via a TTL shutter using the NI-DAQ. However, the response of this laser to being shuttered is very poor. It undergoes significant fluctuations in intensity when turned on and this is true under TTL shuttering as well making it only marginally useful. For future experiments it will be necessary to either replace this laser with a higher quality product or at the very least incorporate an additional mechanical shutter so that the laser does not get turned off.

### 2.1.3 HILO/TIR

### 2.1.4 Azimuthally variable spinning illumination

To establish single molecule detection in *Saccharomyces Cerevisiae* (yeast), it was necessary to compare several types of illumination geometries for widefield microscopes A schematic is shown for epi-illumination[34], TIR[35, 36], and HILO[37] (Figure 2.4). Epi-

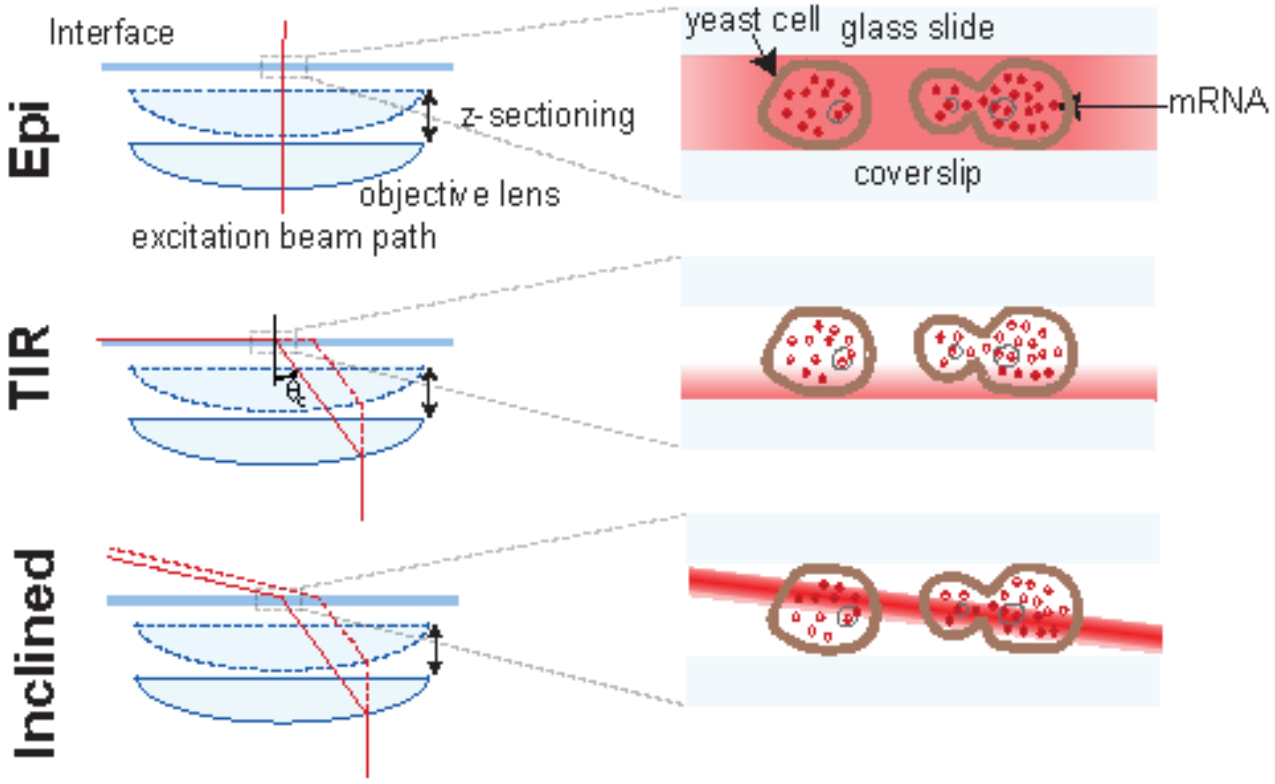


Figure 2.4: Schematic of illumination geometry. Three possible illumination geometries are shown; epi=illumination, total internal reflection (TIR) and Highly Inclined Laminated Optical sheet (HILO). In epi-illumination the excitation path illuminates the entire volume of the sample. In TIR the excitation beam is incident at the critical angle $\theta_c$ and fluorophores are excited by the evanescent wave that does not propagate into the sample, but rather falls off exponentially. For HILO, the beam is incident on the coverslip at a sub-critical angle causing the refracted beam to become thin and highly inclined.

illumination is a common approach to widefield fluorescence microscopy as it only requires that a collimated beam passes through the center of the BFP and illuminates the entire sample volume. This method is suitable for experiments that do not require z-stack images. Since the entire sample is illuminated, one disadvantage is that out of focus fluorophores contribute significantly to the intensity in the image plane leading to poor signal to noise.

TIR is a method where the beam is adjusted to the edge of the BFP until the critical angle $\theta_c$ is achieved. In this setup TIR can be easily accomplished empirically by displacing the beam until the signal disappears and then backing off the adjustment until it returns. At this point, there should also be a reflected beam that can be observed in the excitation path

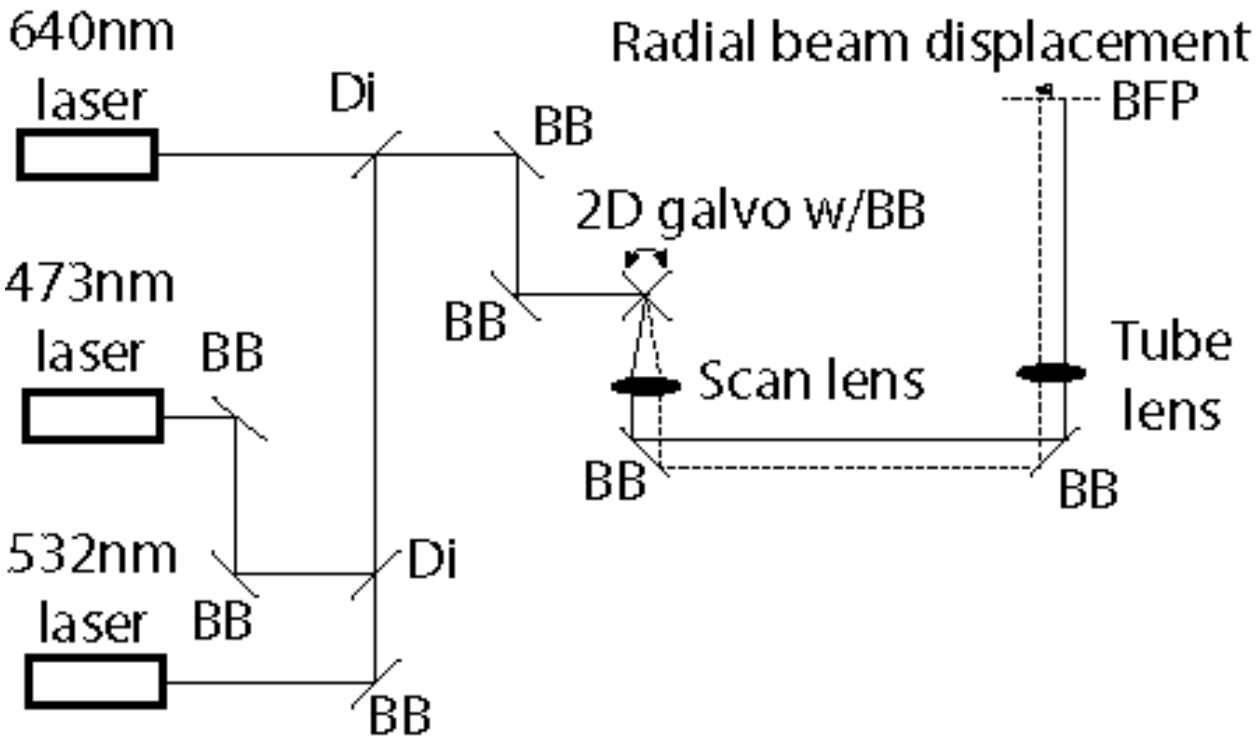


Figure 2.5: Schematic of spinning TIR/HILO setup. Lasers are colinearized in free space and sent into a 2D galvo system. This system is controlled via Matlab code to rotate the beam in a circle. A scan lens converts the spherical focal plane of the beam into a flat focal plane and this is relayed to the BFP.

that disappears as the angle becomes sub-critical. This reflected beam raises challenges in producing high quality images since it is intense enough to appear in the excitation path regardless of emission filters. To address this issue, this beam can be blocked which introduces an astigmatism in the point-spread function of the microscope. TIR is ideal for situations where the signal of interest is located within $100\,\mathrm{nm}$ of the surface. TIR can also be accomplished using small mirrors positioned to reflect the excitation beam at the appropriate radius in the BFP. However, this method is not used in this setup.

While TIR increase signal to noise substantially over epi-illumination, it comes at the cost of z sectioning. In order to accomplish z-sectioning and increased signal to noise in a widefield microscope it is necessary to use HILO. In this illumination geometry, the excitation beam is displaced to a sub-critical angle. The refracted beam becomes steeply inclined

and as a consequence very thinly laminated. This geometry is also frequently referred to in literature as 'near-TIR'. Since this illumination is thin and inclined through the sample it becomes possible to section the sample without exciting the entire sample volume. Since HILO is achieved by adjustment, it is still possible to achieve epi-illuminaion and TIR.

A method to improve the performance of TIR/HILO is to rotate the excitation azimuthally[38, 39]. When the period of the rotation is matched to the exposure time, the fringes in the excitation beam are averaged causing the image to be much more homogeneous. To accomplish the rotation of the beam, the lasers are combined into a colinear path sent into a pair of mirrors rotated by galvonometers (Thorlabs, GVSM002) (Figure 2.5). This 2D-galvo system will rotate the beam based on an analog signal programmed into matlab and sent to the controller via a NI-DAQ analog output. The motor of the mirrors operates with a small angle ($0.2°$) bandwidth of $1\,\mathrm{kHz}$ with a step response time of $300\,\mu\mathrm{s}$ which performs adequately to provide 100 positional updates per period for $100\,\mathrm{ms}$ exposure times. Attempting to drive the mirrors for shorter periods leads to the risk of overheating and temporary operational loss. Therefore, these galvonometers are unsuitable for rapid $1\,\mathrm{ms}$ to $10\,\mathrm{ms}$ acquisitions. An additional potential upgrade is switching to broadband dielectric coating since the frequent humidity and condensation in this particular lab space makes silver protected optical surfaces degrade especially for surfaces pointed upwards as one of the mirrors must.

The rotating illumination is then relayed to the microscope through a telecentric scan lens and a tube lens to the back focal plane (Figure 2.6A). The scan lens is designed using two achromat doublets with an air gap between them. This design is known as a Plössl eyepiece. The performance of such designs has a maximum of $\sim$0.3% F-theta error at a perturbation of $14.3°$[40]. An F-theta lens has a linear response between the incident angle of the illumination and the focal length and the F-theta error is the amount of deviation of the focal surface from a flat plane.. Three equations were used to choose the focal lengths of the lenses in the system where $f_s$, $f_t$ and $f_o$ are the focal lengths of the scan lens, tube

lens and objective lens respectively. $d_1$ is the separation between the 2D-galvo and scan lens. $d_2$ is the separation between the scan lens and tube lens. $d_3$ is the separation between the tube lens and objective lens (Figure 2.6A). $y_o$ and $y_i$ are the beam widths at the objective and initially and are used to determine the size of the illumination in the field of view.

$$\frac{1}{f_s} = \frac{1}{d_1} + \frac{1}{d_2 - f_t} \tag{2.5}$$

$$\frac{1}{f_t} = \frac{1}{d_2 - f_s} + \frac{1}{d_3 - f_o} \tag{2.6}$$

$$\frac{y_o}{y_i} = \frac{d_3 - f_o}{d_2 - f_s} \tag{2.7}$$

When properly aligned, at the halfway point between the mirrors is a conjugate image plane and any dust on the mirrors will be very apparent in the pattern of illumination. To achieve this condition, a light source is placed above the objective to propagate light back through the excitation path and the 2D-galvo system is positioned at the back focal plane of the scan lens. The optimal positioning of the scan lens can be determined by super-sampling the period of rotation. The center of the illumination should be identical to when the system is aligned in epi-illumination and no rotation of the illumination should be observed (Figure 2.6B).

While Ellefsen, et al[38] observe substantial gains in signal to noise in the case of EGFP-tubulin, YFP-STIM1, and mCherry-STIM1 fusion proteins to observe membrane structure and $Ca^{2+}$ ion redistribution. We find that the major gain for a single fluorophore is a much flatter background illumination (Figure 2.6).

Since the laser path can be controlled via computer input it is still possible to achieve epi-illumination and stationary TIR/HILO. An additional consideration is that the mirror position drifts without the motors being supplied with power so that it is necessary to turn the 2D-galvo system on and send a signal to position the beam before each use if spinning

illumination is not used.

### 2.1.5 UV LED

To incorporate the nuclear label DAPI into our imaging scheme, a $375\,\mathrm{nm}$ LED (Thorlabs, M375L4) is added to the illumination pathway. One advantage of this LED is the combined cost including optics, opto-mechanical components, and controller was under $1000. This LED is controlled via TTL through the NI-DAQ. It can also be controlled via analog input. In either case, implementation of a DAQ shutter output through Micromanager was easily accomplished.

The position of the dichroic mirror was chosen for two reasons, the collimated beam attenuates rapidly after about $1\,\mathrm{m}$ and it is preferable that the UV illumination does not rotate since it is hazardous. Furthermore, the broadband dielectric mirrors throughout the illumination path are not coated for UV, which leads to significant loss of power at each reflection.

### 2.1.6 Z-piezo stage

There are two downsides to adjusting the focus by movement of the objective when illuminating the sample in TIR/HILO through the objective. The first is the change in z-position of the objective can effect the collimation of the beam causing it to converge or diverge for large z-adjustments. For the sample sizes similar in thickness to yeast this angular change is negligible.

The second issue is that the change in z-position of the objective changes the location where the beam is incident on the coverslip. This can be directly observed by the z-dependent shift in the distribution of intensity in fluorescent images along the same axis as the beam is displaced in the back focal plane. This cannot be accounted for by moving the sample rather than the objective.

However, first issue can be resolved by fixing the position of the objective and the

regime where this is inconsequential is known as the depth of focus, $DoF$, of the gaussian beam defined as the distance between the two points in the waist of the beam where the diameter is less than $\sqrt{2}$ of the minimum. This can be calculated for the excitation using the diameter of the beam $D$, the working distance of the objective $f_o$, and the wavelength of the laser $\lambda ex$.

$$DoF = \frac{8\lambda_{ex} f_o^2}{\pi D^2} \tag{2.8}$$

For the 100x 1.4NA Olympus objective this depth of focus is $160\,\mathrm{nm}$. Since the sample is approximately $2\,\mu\mathrm{m}$ thick, the objective moves outside the depth of focus to acquire z-stack images.

This piezo stage has been implemented in the microscope, but has not been used for any data collection presented in this dissertation. This stage can be implemented in Micro-manager using the NI-DAQ to send analog output to the driver.

## 2.2 Emission path

There are three emission pathways in the Olympus IX-82 frame. The first directs the emission to the eyepiece for direct observation. The second and third pathways can be switched by manually adjusting a mirror in the frame. The third pathway has only been used for back focal plane imaging, confocal imaging witha Yokogawa CSU-10 confocal unit, and for comparison for demo cameras.

One utility of the third path is that the 1:1 image plane is formed on the CCD mounted to the camera frame so that it can be used to align a camera that has been moved away from this plane to incorporate additional optics. The correct focal alignment is chosen when the camera being aligned is parfocal with the camera mounted on the microscope frame.

### 2.2.1 Single Channel

To perform the majority of experiments presented in this dissertation a single channel emission pathway was used. This pathway was modified from the default configuration (Figure 2.9) by displacing the camera away from the image plane and creating a conjugate image plane using a relay lens to also add additional magnification where magnification $M$ is the ratio between the distance between the relay lens and the aperture, $a$ and the relay lens and the camera, $b$.

$$M = \frac{b}{a} \tag{2.9}$$

The need for additional magnification was driven by the requirement to properly sample the fluorescent intensity of single fluorophores for accurate spatial localization of their intensity. The optimal sampling for a widefield microscope based on the Nyquist bandwidth is done at a rate that is twice as frequent as the Abbe diffraction limit $d_{res}$ and is dependent on the wavelength of the emission $\lambda_{em}$ and the numerical aperture of the objective $NA$[41, 42].

$$d_{res} = \frac{2\pi}{\Delta k_{xy}} = \frac{\lambda_{em}}{2NA} \tag{2.10}$$

The $NA$ is dependent on the index of refraction $n$ and the half aperture angle of the objective $\theta_{1/2}$

$$NA = n \sin \theta_{1/2} \tag{2.11}$$

Therefore, to optimally sample the pixel size $Pixel_{xy}$ is half of $d_{res}$. The equation for $Pixel_z$ depends of the same parameters, but the optimal z-sampling is larger than xy-sampling[43].

$$Pixel_z = \frac{2\pi}{\Delta K_z} = \frac{d_{res} \sin_{1/2}}{1 - \cos \theta_{1/2}} = \frac{\lambda_{em}}{2n(1 - \cos \theta_{1/2})} \tag{2.12}$$

### 2.2.2 FRET

To image molecules undergoing Fluorescence Resonant Energy Transfer (FRET)[44], a bandpass filter with two bands compatible with Cy3 and Cy5 is placed in the filter cube without an emission filter. In the emission path past the tube lens and aperture, a dichroic mirror to separate emissions from Cy3 and Cy5 is placed at the image plane (Figure 2.10.

The Cy5 emissions are transmitted and follow the same path as in the single channel case with additional magnification. The Cy3 emissions are reflected and an additional path of the same magnification is constructed with an additional relay lens and dichroic mirror to recombine the paths. The aperture is then adjusted and the emissions are diverted to cover only one half of the CCD each. Emission filters for Cy3 and Cy5 are placed before the relay lenses and after splitting the path for each path.

Alignment of these two channels is carried out by using a probe designed to be in an intermediate FRET state so that spots can be observed in both channels. These spots are then used to calculate an affine transformation to map one channel onto the other. In addition, Cy3 and Cy5 are measured separately so that bleed through can be accounted for particularly in cases where FRET efficiency is not designed to be maximum. Conveniently there are no dynamic FRET experiments carried out so that sensitized emission[45] is less important than the ability to treat FRET as a binary marker of colocalization of the two fluorophores on a target RNA.

### 2.2.3 Fluorescence anisotropy

Fluorescence anisotropy[46, 47, 48] is a technique where the polarization of emission is separated and the intensities are compared to determine if the emission is correlated to the polarization of the excitation. If the timescale of the rotation is faster than the timescale of photon emission then the resulting emission is scrambled and there is no anisotropy.

If instead the orientation of the molecule is fixed during the timescale of emission, then the resulting emission is correlated to they polarization of the excitation and anisotropy

is observed. This can be calculated by the ratio $r$ of the difference in intensity of each polarization $I_p$ and $I_s$ to the total intensity $I_{total}$.

$$r = \frac{I_p - I_s}{I_p + 2I_s} = \frac{I_p - I_s}{I_{total}} \tag{2.13}$$

The alignment of this system is particularly complicated since it is preferable to avoid any magnification difference in the three paths. A linear polarizer is placed in the excitation path to ensure that the laser light is limited to one polarization. The emission path is similar in design to the FRET path with the addition of a pair of polarizing beam splitter cubes. These reflect s-polarized light and transmit p-polarized light. Each of these cubes is mounted on a kinematic platform that allows 2-axis adjustment of the path of the s-polarized light. The aperture is adjusted so that three channels can fit on the CCD. Since it is also desirable to maintain the proper sampling, there are tight spatial constraints to the alignment that make it difficult to make the third path the same length as the two paths separated by the polarizing beam splitters.

While this configuration allows for FRET and anisotropy to be measured, blocking the Cy3 path and only observing the polarization simplifies the alignment of the optics and allows for fluorophores other than Cy5 to be observed.

### 2.2.4 Back focal plane

The primary motivation for imaging the back focal plane (BFP) was to understand the performance and alignment of the spinning TIR/HILO system. This imaging design provides information about the location of the beam and how much of a period is completed per exposure and represents the Fourier transformation of the image[49]. Several advanced microscopy techniques involve placing diffraction gratings at the BFP to separate different focal planes of information into unique emission paths.

The laser spot can be observed in the center of the BFP. When the laser is misaligned

the spot will be elliptical rather than circular. Furthermore, a perturbation of the spot away from the center reveals two reflected spots. These can also be used to determine if the beam propagates vertically or is tilted.

Some microscopes come equipped with a Bertrand lens, which switches the image from real space to Fourier space image. However, this is not part of this microscope frame. To achieve BFP imaging, the location of the imaginary image of the BFP is calculated to be several meters away from the tube lens since it is approximately $0.2\,\mathrm{cm}$ inside the focal point of the tube lens. This means the light is nearly collimated coming out of the tube lens. A $10\,\mathrm{cm}$ lens is placed as close to the tube lens as the microscope frame would allow and an image of the BFP plane is formed on the CCD at the focal point of this second lens. Since this type of imaging is done as a reference for alignment quality a second camera was placed on an optical post at this position and the primary camera is left in position for data acquisition. The image cannot be recorded simultaneously since the BFP imaging is accomplished using the third path in the microscope frame and these paths are switched manually via a mirror(Figure 2.12).

An improvement to this setup would be to move the camera to the second image path and pick off some of the emission so that the image plane and BFP can be simultaneously recorded.

### 2.2.5 Astigmatism based tracking configuration

Much of the effort in this microscopic system is done to obtain the best image possible, which inherently means that astigmatisms are to be avoided. However, one useful property of an astigmatism is the z-information it carries.

Since an astigmatism is asymmetrical in the z-axis, introducing an astigmatism such as blocking the reflected beam in the stationary TIR/HILO setup causes a change to in the intensity distribution due to the missing information. Depending on the shape of the object blocking the beam, part or all of one side of the PSF can be missing. For example

using a flat edge to block the reflected beam can cause a semicircular airy ring around the center spot. Therefore, for high quality images it is desirable to minimize the amount of information lost to blocking the beam.

However, another type of astigmatism can be introduced which does not remove information but causes one axis of the emission to focus more rapidly than the other axis (Figure 2.13. Complicated astigmatisms can be designed to provide z information by the addition of otpics at the BFP[50]. In this design, a cylindrical lens is placed just past the image plane formed by the tube lens. In the case where the lens is placed at the image plane, there is no change in the image. The resulting image is at the proper z-position when the x and y intensity patterns are equally out of focus. This happens at the midpoint between the two focal points. There are two consequences of this astigmatism. The intensity rapidly diminishes when the object is out of focus in either direction making out of focus spots much less apparent. The second consequence is that the PSF has multiple lobes rather than Airy discs around it creating difficulty in peak detection as a means of locating the spot. While the alignment of the cylindrical lens is fairly simple, the criteria for selecting the position is less straight forward since it is dependent on the computational constraints of tracking. A good choice for the location of the lens is therefore one that provides decent contrast between the image in each z-step while not allowing the astigmatism to stretch the image at the desired z-limits of tracking outside of a reasonable boundary such as a 9x9 pixel window. This is to allow for updating the z position via computation within the timescale of diffusion.

### 2.2.6 Computer control

All of the data presented was acquired with an EMCCD camera (Andor, iXon+ 897) with few exceptions which will be mentioned explicitly. Control of this camera was accomplished using Micromanager although some time was spent with modifying in-house code to work with this system. Micromanager control creates a configuration file for the devices

it initializes. Certain properties which would be convenient to be set upon launch are not actually included. Custom code was added to the configuration file to set the camera to Frame Transfer mode, set the temperature setpoint of the camera, enable the focus wheel on the microscope frame, and activate the transmitted lamp. Further time was spent debugging aging device drivers with Micromanager and it was determined that the AOTF was not capable of receiving serial commands at the rate coded into Micromanager, which was a resolved through a version patch fixing the code error we located. The majority of devices are controlled via serial port with the exception of those that are controlled via analog or TTL output via an NI-DAQ which can be controlled via Micromanager.

In particular the 2D-galvo system was initially controlled via a script in Micromanager. This was performance limited by the large amount of memory that Micromanager consumes. It was observed that the signal causing the beam would occasionally pause when controlled via Micromanager. This led to switching to Matlab control which has built in functions to work with the NI-DAQ.

### 2.2.7 Camera settings

A variety of camera settings were employed depending primarily on the resultant image quality. Initially, with the gas laser system providing low excitation power, one to two second acquisitions were performed with the EMCCD gain set between 500 to the maximum of 1000. These settings were chosen to strech the brightness of the image across the entire dynamic range without saturation based on the brightest sample in a set of experiments and then propagated to all other samples. With improvements to the illumination scheme, it was possible to reduce the gain and the exposure time so that images are acquired with between 5 and 150 gain and at $100\,\mathrm{ms}$. In principal EM gain is used to increase signal to noise, however, with stronger excitation intensity more photons are produced and EM gain becomes less necessary to achieve good signal to noise. It is desirable to keep exposure times to a minimum to decrease acquisition time and laser power to a minimum to

increase fluorescence lifetime. Therefore once signal quality is adequate, any change in signal intensity is compensated with a change in gain rather than the other two parameters.

Frame rates much faster than that are used for some experiments. In these cases, the frame time is dependent on the vertical and horizontal voltages and clock speeds, the binning, and the region of interest. Frame transfer mode is always selected to be on, which allows the camera to continue acquisition while the data is being read from a secondary set of pixels. The effect of cropping on frame time is more significant on the vertical axis so that a 512 by N cropped strip will perform at nearly the same frame rate as a N by N cropped array. This is due to the vertical cropping affecting the size of the image array and not the serial register. This cropping is not as effective as cropped mode which requires the addition of an aperture to mask the unused pixels from light. In all cases cropping simply means that the values outside the region of interest are disregarded at readout thereby increasing acquisition rate.

## 2.3 Measuring the Point Spread Function

The PSF of a microscope defines how the microscope transforms a point source into the observed image. Measuring the PSF is important since it allows for a better understanding of how objects in the sample will appear since every object can be interpreted as a superposition of many point sources. Mathematically, the convolution of an object with the PSF reults in the observed image. If this is fourier transformed then the image is simply the product of the fourier transform of the object and the fourier transform of the PSF also known as the optical transfer function.

The motivation for measuring the PSF is therefore to understand how diffraction and spherical aberration blur the object as it is transformed into the image and to be able to use that knowledge to reconstruct the true image from the observed image. This would naturally seem to be to simply perform the inverse operation in fourier space, which is known as deconvolution and will be discussed in the following chapter.

A further use for the PSF is to understand the inherent width of a spot and how to model a spot when performing image analysis whether or not deconvolution is performed.

To measure the PSF, Cy3 or Cy5 labeled DNA oligos are placed in oxygen scavenging buffer on a coverslip and mounted on a slide. These oligos interact non-specifically with the glass surface and some portion of them stick. Z-stack images are acquired of spots at the surface. Since these molecules are smaller than the diffraction limit, the intensity profile represents the PSF of the microscope. This is dependent on the wavelength of emission so that shorter wavelength fluorophores will produce higher resolution images. Therefore, the optimal sampling is always chosen based on the shortest wavelength used.

The image can be fit as the intensity $I(\theta)$ of the Fraunhofer diffraction pattern of a circular aperture using the Bessel function of the first kind $J$, the max intensity $I_0$, the wavenumber $k$, the radius of the aperture $r_a$ and the angle of observation $\theta_o$.

$$I(\theta_o) = I_0(\frac{2J_1(ka\sin\theta_o)}{ka\sin\theta_o})^2 \tag{2.14}$$

The radius of a spot is then defined as the FWHM of this intensity distribution or approximately 1.5 pixels.

## 2.4 Confocal imaging

To perform confocal imaging a Yokogawa CSU-10 is mounted to the microscope frame via c-mount. In this configuration, the excitation must be coupled to a fiber and connected to the side of the confocal unit. Additional magnification is not practical in this setup since by design the unit needs to be properly positioned for both excitation and emission which is only feasible when placed at the image plane formed by the tube lens. This confocal unit includes a nipkow disk which was partially damaged lent to us by courtesy of Prof. Yoda in the Mechanical Engineering department. The extent of the damage degrades the image quality of more than half of the field of view. The nipkow disks have a microlens array

that patterns the illumination and a pinhole array that restricts out of focus illumination from contributing to the image intensity pattern. A dichroic between the two discs allows the excitation beam to be transmitted and reflects the emissions onto the CCD. As the disk spins the pinholes sweep the entire area of the CCD during an exposure. In this model, the rate of the rotation is fixed. The EMCCD camera was mounted to the confocal unit.

Additionally, a Zeiss scanning confocal microscope was tested with an RNA-FISH sample. However, the sensitivity of the photo multiplier tube was not at a level to detect single fluorophores.

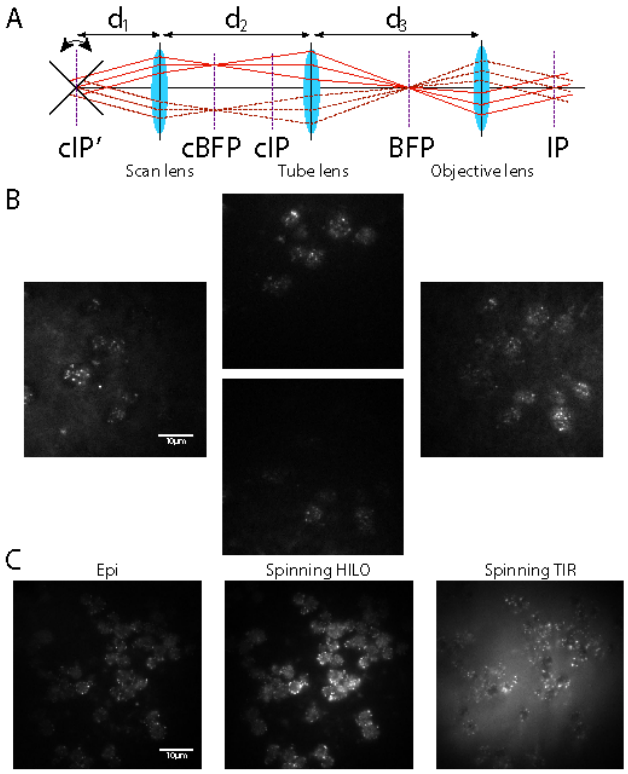


Figure 2.6: Alignment of spinning illumination. A. Ray diagram of two points in the cycle. The illumination when properly aligned is focused at the BFP and conjugate BFP. The illumination should be collimated out of the objective and the illuminated spot should not rotate. B. Misalignment of the scan lens. When the system is poorly aligned, the illumination either diverges in the case of misalignment of the 2D galvo system with the scan lens or it converges if the scan lens is too close to the tube lens. In the second case the spot of excitation rotates in the field of view. C. Comparison between epi-illumination and spinning versions of TIR/HILO. Epi-illumination produces a dimmer and more localized region of excited fluorophores than either spinning TIR or HILO. In TIR the spinning illumination produces a more even region of excitation compared to stationary TIR, but suffers from the inclusion of the reflected beam in the background. In comparison to spinning TIR, spinning HILO seems to have less issue with contrast and still retains the more uniform excitation.

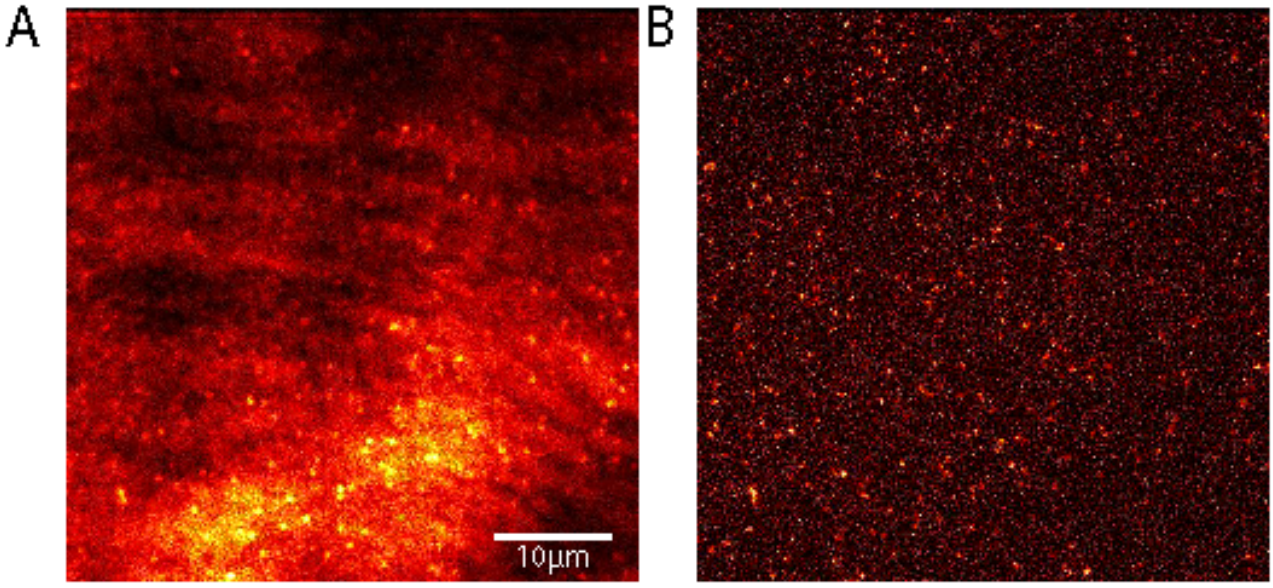


Figure 2.7: Spinning TIR on single fluorophores. The left image shows an image of TIR illumination of single fluorophores. The right image shows the same field of view at the same angle of incidence of the laser but the illumination is now rotated for one complete period per exposure.

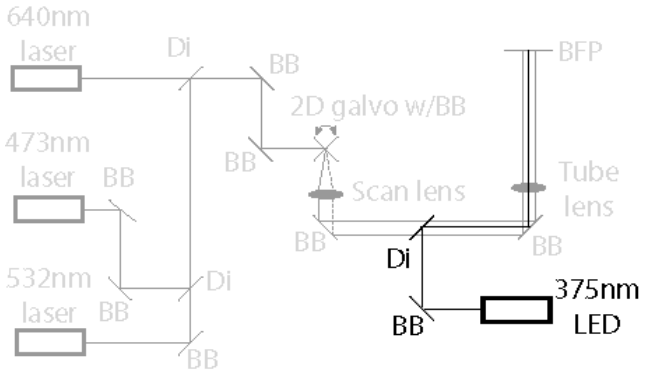


Figure 2.8: Schematic of UV setup. A $375\,\mathrm{nm}$ LED is added to the spinning system by collimating the LED with a $5\,\mathrm{cm}$ lens and inserting a dichroic mirror between the two mirrors relaying the spinning illumination to the BFP.

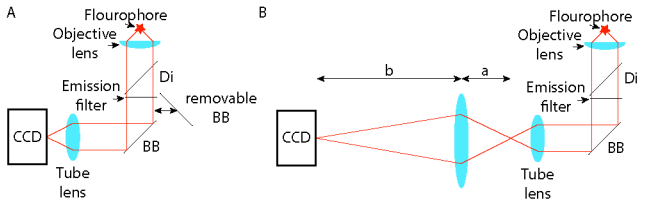


Figure 2.9: Schematic of single channel emission path. A. Mounted via c-mount. The camera can be directly attached to the frame of the microscope. In this configuration there are no additional lenses or filters besides those inside the microscope frame. B. Camera configuration with additional magnification. The camera is supported on optical posts and the image is formed via a relay lens. This allows for additional optics such as filters to be placed in the optical pathway and additional magnification.

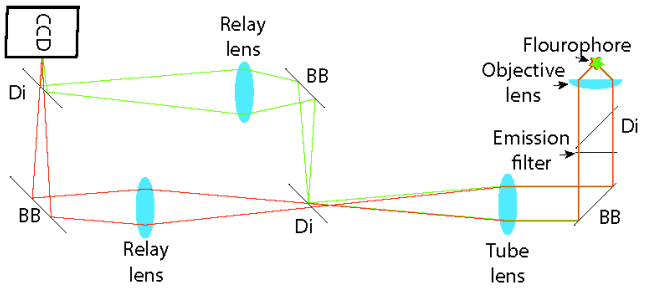


Figure 2.10: Schematic of a dual channel configuration. This configuration places a adjustable mechanical slit at the image plane. Two paths are created of equal length by a dichroic mirror splitting the illumination. Both paths are focused onto the CCD and are adjusted to each cover one half of the CCD. A dual-band bandpass dichroic mirror is in the filter wheel and emission filters are placed in the optical path after the dichroic that separates the paths.

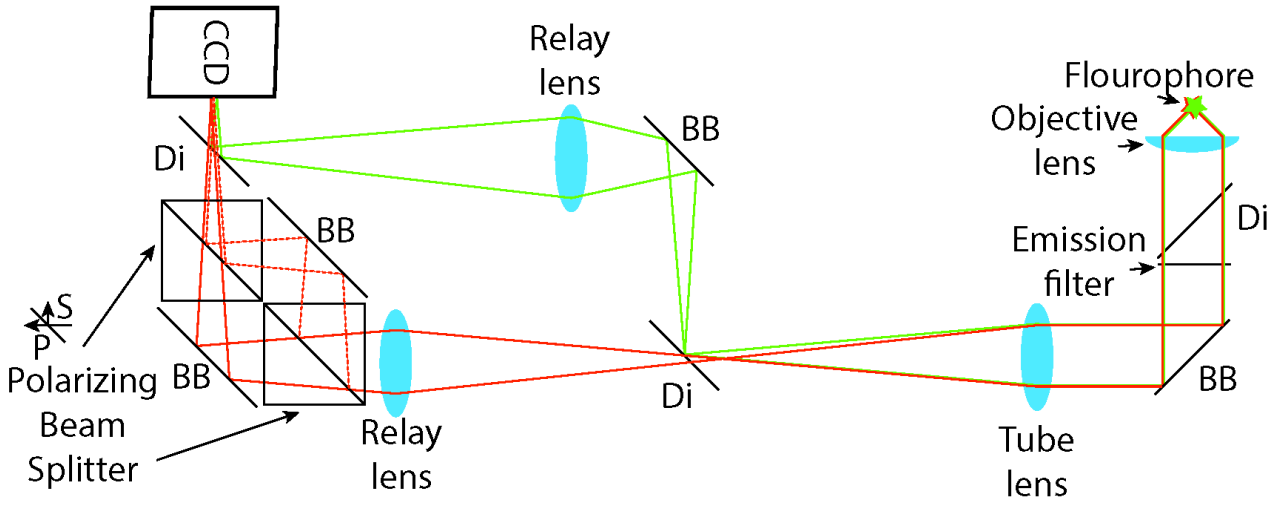


Figure 2.11: Schematic to observe fluorescence anisotropy. This schematic is similar to the dual-channel pathway with the addition of two polarizing beam splitter cubes in the primary path. These cubes separate the electromagnetic field into the p and s polarizations, e.g. parallel and perpendicular polarizations. The s-polarization is the reflected beam and requires standardization to accurately measure anisotropy due to the transmission efficiency of p-polarized light being higher than the reflection of s-polarized light.

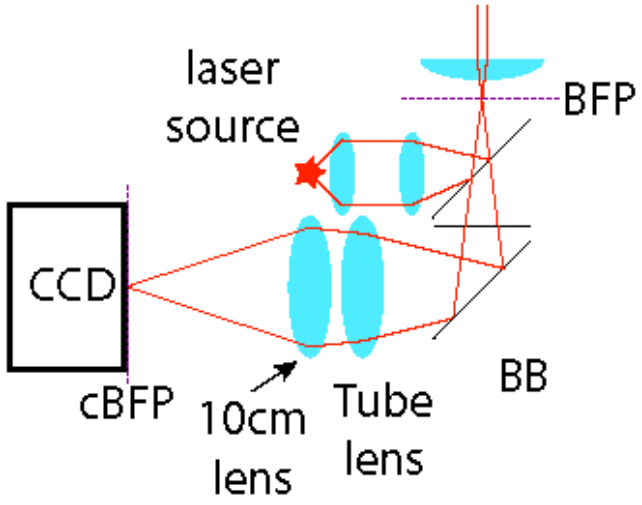


Figure 2.12: Schematic to observe the back focal plane. The camera is switched to the third optical path in this design. A $10\,\mathrm{cm}$ lens is placed directly after the tube lens. The tube lens focal length is $18\,\mathrm{cm}$ and the location of the back focal plane is just inside the focal point of the tube lens causing the BFP illumination to diverge.

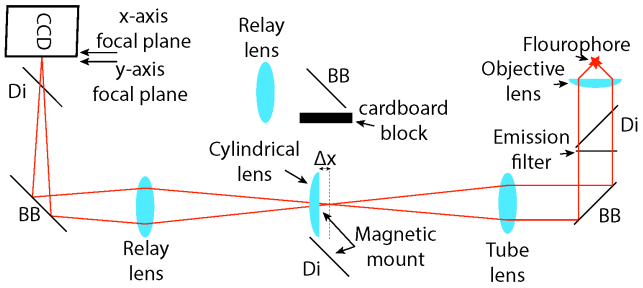


Figure 2.13: Schematic to introduce an astigmatism. In this design, the single channel path with magnification is modified so that a cylindrical lens is placed close to the image plane. This lens introduces an astigmatism where the light focuses at different axial positions along the light path along the x and y axes of the cylindrical lens. This lens is mounted in a rotating lens mount so that the x and y axes of the lens can be matched to the axes of the camera image. The optics are now mounted on removable magnetic bases so that the emission path can be quickly and accurately switched between several configurations.

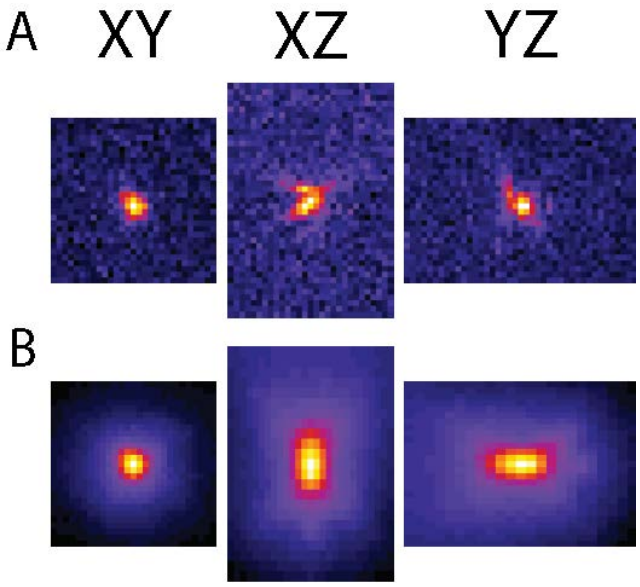


Figure 2.14: Measured Point Spread Function. A. An image of a single Cy3 with xy, xz, and yz projections is shown. B. The PSF of the microscope is estimated as the average of many Cy3 excited at $532\,\mathrm{nm}$.

# CHAPTER 3

# IMAGE PROCESSING AND ANALYSIS

Image analysis is typically performed for two reasons. The first and primary reason being to extract quantitative information about a signal in the image and the second being to enhance the image to be able to perform the first function. In general it is a bad practice and most often a frustrating waste of time to attempt to produce good images out of bad images using image enhancement. Therefore, it is critical to establish a set of criterion for which to proceed with analysis. Since the signal that is being detected is in general Cy3 or Cy5 throughout, it can be assumed that a set of criteria for a good analysis pipeline for diffraction limited spots is most important.

Another important set of design criteria include those for image analysis techniques that can detect single cells and organelles such as the nucleus automatically. There are some universal criteria.

- There should be a minimum of user input.
- There should be no subjective thresholds or quality parameters.
- Image enhancement should be unnecessary or unambiguous.
- Computation should be rapid.

The following image processing techniques are going to be discussed with the raw data in Figure 3.1 as a reference. This data is a typical set of FISH image acquisitions and in particular the TLC1 RNA tends to localize to the nucleus. This leads to spots that are not sparsely distributed and therefore harder to separate during detection.

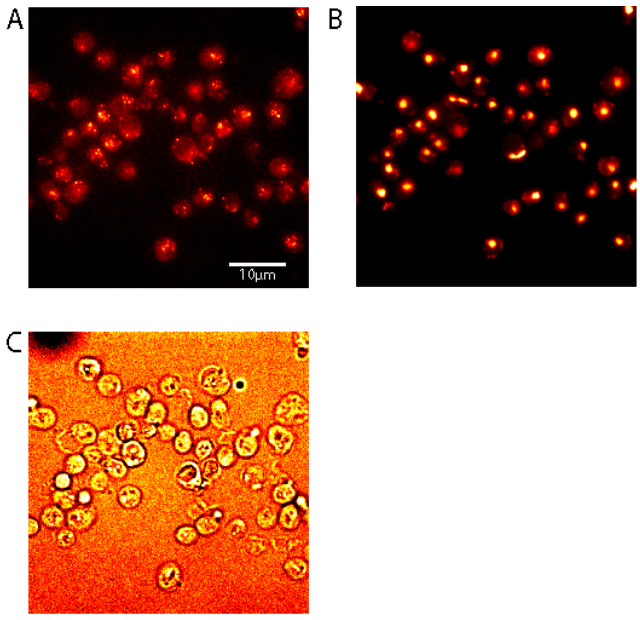


Figure 3.1: Raw data. A. Cy3 channel. The center slice of a channel representing TLC1 RNA spots is shown. B. DAPI channel, The center slice of the channel representing the nucleus is shown. C. DIC. Cell boundaries are chosen based on a DIC image which is shown as the center slice.

## 3.1 Analysis methods available from literature

### 3.1.1 FISHquant

FISHquant is a Matlab based application that performs every step necessary to extract spot counts from RNA FISH data[51].

The workflow of FISHquant is to load in an image stack and enhance the image using background subtraction where the background is modeled as the image smoothed by a wide gaussian filter. Then the image is filtered for noise using a narrow gaussian filter matched to the width of a spot. Gaussian filters function as low pass filters so that spatial variations that are wider than the filter are unaffected and high frequency variations are smoothed to conform to the shape of the kernel of the filter. The program creates a maximum projection of the fluorescent intensity data with which to perform cell segmentation using ROI processing functions from the image processing toolbox of Matlab (Figure 3.2. Transcription

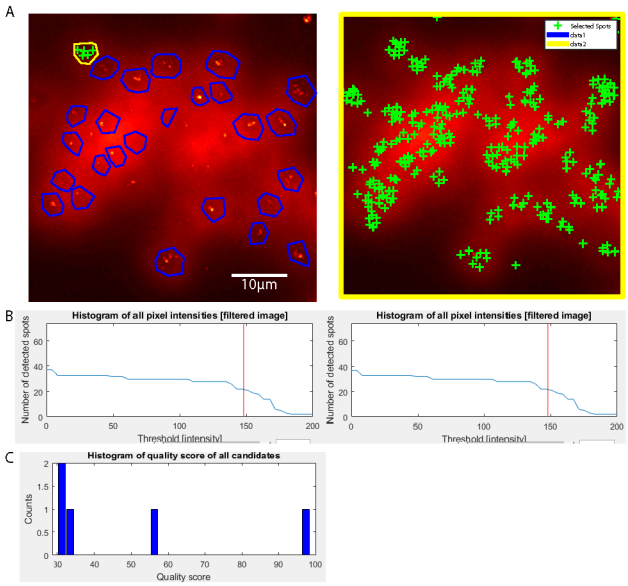


Figure 3.2: Fishquant boundary selection and filtering. A. Boudary selection. Cell boundaries are drawn by hand on a smoothed and background subtracted maximum projection of the image of Cy3 image (left). If cell boundaries are not selected and spot detection is performed the entire image is selected as the region of interest (right). B. Intensity thresholding. The first layer of parameter choice is the intensity threshold of the first cell. C. Quality score. Spots are also further filtered by choosing boundaries of a quality score.

sites and nuclei can also be selected by hand in the segmentation workflow.

After the image is segmented, the parameters for spot detection are chosen. The first level of selection is intensity based and is done in a similar way to other methods by plotting the number of spots detected at each cutoff and displaying the number of spots in an image for manual inspection and selection. One flaw in this is design is that it does not treat the whole image unless you skip the cell segmentation step, which could introduce bias due to an non-uniform illumination. This can be avoided by skipping the order (they are numbered in the GUI) and returning to cell segmentation after spot identification. Additionally, since images can still be non-uniform in illumination after a simple 2-dimensional Gaussian approximation of background, a local choice of thresholding or a better way to flatten the image is desirable. After the threshold is chosen the projections of the image in xy and xz can be inspected to understand the choice of spots (Figure 3.3).

After the correct intensity is chosen. the spots are each fit with a Gaussian intensity

profile that requires the choice of upper and lower limits for seven more arbitrary parameters that are then abstracted into a quality score that also has manually selected upper and lower limits (Figure 3.5). After this point, fitting is applied to all spots in the image inside of segmented boundaries. The program can then display spot locations over an intensity image (Figure 3.6

The fit parameters of the 3-dimensional Gaussian are $175\,\mathrm{nm}$ for $\sigma_{XY}$ and $477\,\mathrm{nm}$ for $\sigma_Z$. The background intensity is estimated to be 466 and the amplitude of the Gaussian is estimated to be 622 by the fit. These parameters are similar to fits by other methods used later.

### 3.1.2 Star Search

One method to detect spots is to separate them from the background using an intensity threshold. In this java applet, the image is loaded and a maximum projection is shown. This follows the same procedure as the Matlab code made available through the Raj lab[52]. In this code, fluorescence intensity images are maximum projected and spots are converted to a binary image using a threshold which is chosen by inspection. Ideally, there should be some threshold where the number of spots is stationary over some range of intensity. Spots are then labeled and counted using the Matlab function bwlabeln().

The biggest flaw in this approach is that it assumes that all spots are sparsely distributed. In the case where this is true the method is quite effective and rapid. Furthermore, even though the applet uses a maximum projection there is no need to limit this to a maximum projection when the spots are adequately sparse. The choice to use this method is therefore dictated by the subcellular distribution of the spots and the density.

### 3.1.3 Other spot detection algorithms

Other means to detect spots include nonlinear fitting of detected peaks with a Gaussian. This can either be done via nonlinear regression using built in Matlab functions or using a

simplified least squares Gaussian mask approach[53]. The method used by Zenklusen, et al[28] involved hand drawn cell borders, thresholding of the DAPI channel, and Gaussian fitting of spots in a 2-dimensional maximum projection. Nascent transcripts were fit with a sum of Gaussians.

## 3.2 Image Enhancement

The initial attempts at measuring Cy5 inside of cells were done at low laser power and high gain. Even with EM gain, the image quality was poor and various methods to enhance the image quality were explored. The majority of these methods were explored before any improvement was made to the microscope. After the image quality was improved through changes to equipment and alignment, many of these techniques became unnecessary. Image enhancement also can be computationally intensive making image analysis take significantly longer.

### 3.2.1 Linear filtration

One of the most basic techniques in signal processing is linear filtration. In the case of 2-dimensional images linear filters replace the center pixel with the sum of pixels within some window based on some combination of weights. An important consideration with linear filters such as the averaging, Gaussian, or Laplacian of Gaussian (LoG) filters, is that objects smaller than the defining width of the filter will become shaped like the Gaussian kernal of the filter. This means that all of the noise in images filtered with a narrow Gaussian filter ends up also looking like a spot albeit of much less intensity.

The Gaussian filter is a popular choice to smooth an image to reduce the contribution of noise. Features that are smaller than the specified width of the Gaussian, $\sigma$, are blurred so that a good choice of parameter is to match the width of the gaussian to the diameter of the PSF of the microscope (Figure 3.8B). A wider Gaussian filter can be used to model the background intensity of the image (Figure 3.8C) and is subtracted from the smoothed

image. This subtraction can also be weighted so that the image is less truncated around the peaks. An alternative is to just take the average of all the pixels in some window to represent the background. In this case, the kernal of the image filter produces square artifacts which can be more perceptible depending on the size of the filter used (Figure 3.8D).

The performance of a linear filter on a 3-dimensional image can be improved by concatenating the filter into a $(mxnx3$ matrix to give some consideration to neighboring focal planes. The reulting 3-dimensional filter is normalized This filter still functions the same way as the previous filter, but the enhancement of the spots is stronger when they appear in multiple frames..

### 3.2.2 Binary morphological operations

Binary morphological operations are incredibly useful for segmentation of an image or for reducing noise in selected objects in an image among other uses. The first step to implement binary morphological operations is to convert the image from grayscale to binary using an intensity threshold or in the case of edge detection a Sobel filter.

With this binary image erosion, dilation, or compound operations such as opening or closing can enhance the ability to detect objects. The process of erosion is useful to clear noise from detected borders. Often, single pixels will have poor connectivity along the border. A structuring element is defined such as a disk of radius two. This structuring element is passed across the image and if the this fits into the image then the value of the center pixel in the result is one and is zero otherwise. Essentially, the dot product of the structuring element at some pixel with index $(i, j)$ is equal to the area of the structuring element. Dilation does the opposite of erosion and the pixel at $(i, j)$ is one if the dot product of the structuring element with the sub image is non-zero. In the case of the disk, erosion would reduce the border by the radius whereas dilation would expand the border by the radius.

Another binary concept to consider is connectivity. The borders around an object can

be cleaned by applying a threshold of connectivity in either 2D or 3D. In this image pixels with value of 1 would have pixels that are over the threshold in all 4 or 8 neighboring pixels for 2D or 6, 18, or 26 neighboring pixels for 3D.

One of the most useful functions in Matlab for these types of images is bwlabeln(), which counts the number of connected objects in the image and creates an equivalent array where these objects have their value set to the label. This allows for easy counting of the objects in the image or for the use of the binary image as a mask to perform operations on one object at a time.

### 3.2.3 Top-hat transformation

The top-hat transformation turns all objects that the structuring element fits inside to zero. This means that only objects that have intensity variation that is smaller than the structuring element survive.

Since the shape of spots is approximately Gaussian. This suggests that a good structuring element would be a Gaussian that is determined by the width of the spot or slightly larger. The resulting image should in principal remove all of the background and only return high frequency variations such as noise (Figure 3.10A). When the structuring element is increased in size to the radius of the spot, noise becomes less significant and signal becomes more obvious (Figure 3.10B). The top-hat transformation performs fairly well to reduce the background and improve contrast between the signal and background when the spots are uniformly illuminated. However, this method struggles to enhance spots that are dimmer. The spots become quite obvious when the structuing element is the same size (Figure 3.10C) or larger than the spots (Figure 3.10D).

### 3.2.4 Fourier transformation

Since the imaging process is represented by the convolution of the true image with the PSF of the microscope in real space, it would make sense that image enhancement would be

easier by performing the Fourier transformation to be able to simply divide the image by the optical tranfer function, which is the Fourier transformation of the PSF.

Additionally, since noise appears in the spectrum at high frequencies and the background appears at low frequencies, the Fourier transform of the image can be filtered for these frequencies to improve the signal to noise.

### 3.2.5 Blind deconvolution

Blind deconvolution is the process of simultaneously reconstructing both the true image $O$ and the point spread function $PSF$ in the case that both are unknown. This can also be used to reconstruct the degradation process $n$.

$$O = I * PSF + n \tag{3.1}$$

There are a variety of approaches to inverting this process[54] and addressing all of them would be challenging. The algorithm in the Matlab image processing toolbox takes a Maximum Likelihood approach (Figure 3.12.

### 3.2.6 Lucy-Richardson deconvolution

The Lucy-Richardson algorithm for deconvolution was developed from an application of Bayes' theorem to images[55] or radio-astronomical observations[56]. This algorithm takes the image formation process where the observation $O$ is the convolution of the true image $I$ and the point spread function $PSF$. Unlike blind deconvolution, the $PSF$ is assumed to be known in the Lucy-Richardson deconvolution algorithm.

This is a straight forward application of Bayes' theorem where the probability of the true image $I$ is conditional on the observation $O$.

$$P(I_i|O_k) = \frac{P(O_k|I_i)P(I_i)}{\sum_i P(O_k|I_j)P(I_j)} \tag{3.2}$$

With some substitutions this can be turned into an iterative procedure that requires an estimation of the true image at the first step.

$$P_{t+1}(O_i) = P_t(O_i) \sum_k \frac{P(O_k|I_i)P(O_k)}{\sum_j P(O_k|I_j)P_t(I_j)} \tag{3.3}$$

The initial estimation frequently used is Bayes' postulate which assumes that all prior estimations of the true image are equally probable. The result of applying this algorithm, which is contained in Matlab's image processing toolbox can be seen in Figure 3.13. This function requires the input image and the $PSF$ to obtain the reconstructed image.

### 3.2.7 Nearest-neighbor deconvolution

Nearest-neighbor deconvolution is not a deconvolution technique in the sense that it performs the mathematical inverse of the image formation. However, it is called deconvolution because in a broad sense because the purpose of of this type of algorithm is to remove blur from out of focus emission from the image.

The nearest neighbor method assumes that the out of focus emissions detected in the plane of focus $I_{n_observed}$ is primarily due to blurring of the images immediately above and below the plane of focus, $I_{n\pm1}$. The blurred images, $\nu_{n\pm1}I_{n\pm1}$, are calculated by applying a Gaussian filter to the images before subtraction.

$$I_{n_observed} = I_n - \nu_{n+1}I_{n+1} - \nu_{n-1}I_{n-1} \tag{3.4}$$

This method is a decent way to improve contrast by removing the background contributed by fluorophores in neighboring planes. However, this method also requires that optimal sampling in Z be performed due to needing the neighboring planes to have correlation to the plane of focus. If the image is under sampled, then the images do not represent the blur in the plane of focus.

### 3.2.8 Image entropy

The entropy in an image can be used to estimate the quality of the image and the distance from focus [57]. Shannon's entropy is a concept from information theory where the more detailed the information the higher the entropy. In the context of an image, the more out of focus the image becomes the more homogeneous the image becomes due to diffraction. This blur contains less information about the location of the object and therefore has less entropy. Another way to consider this is that as entropy increases the disorder of the image increases meaning that the intensity of a pixel becomes less correlated to its neighboring pixels.

Since there is little gained to analyzing images that are out of focus, image entropy was treated as a possible selection criteria for images that were worth computation time. To compute the quality of the image the variance in the image slice is calculated and then a cutoff is chosen that represents the point where the cells become out of focus (Figure 3.15.

### 3.2.9 CIDRE

In most optical setups the illumination is uneven and substantial effort must be taken to achieve the best quality image. To reduce bias and correct for the uneven illumination there are methods that use a reference image and there are methods that perform the correction in retrospect without any reference. One such method is Corrected Intensity Distributions using Regularized Emergy minimization (CIDRE)[58]. This method solves the inverse image formation process using an additive term $z$ and a gain term $v$ to obtain the image $I$ from the observation $I_o$.

$$I(x) = \frac{I_o(x) - z(x)}{v(x)} \tag{3.5}$$

This problem is made challenging because it is difficult to know a priori how the signal attenuates or how shot noise and other additive contributions that are independent of the signal will effect the observation. In principal, a retrospective method is good because

it does not require any particular additional characterization of the imaging system. The method of CIDRE assumes that the distribution of intensities of a single pixel can be related to an underlying distribution of intensities that is common to all the pixels in the CCD. The result of CIDRE processing can be seen for the example images (Figure 3.16).

#### 3.2.10 CLEAN algorithm

The CLEAN algorithm was introduced in astrophysics to deconvolve radio signals. In this method the observation is assumed to be a 'dirty' signal on a 'dirty' map of background intensities. The 'dirty' map is then 'cleaned' by the iterative removal of the peaks in the image that are represented as some 'clean' signal that has been convolved with the point spread function to produce the 'dirty' signal. At the end of the process the peaks detected by the algorithm are convolved with the PSF and added back to the 'clean' map to reconstruct the image.

In the context of FISH this is an ideal process to use when the spots are clustered. In this case, intensity thresholding would turn these spots into an amorphous blob and deconvolution by other algorithms would struggle to separate overlapping spots. The removal and detection makes each subsequent spot easier to identify.

### 3.3 GPU acceleration

Matlab image processing functions are largely gpu enabled. Image filtration can be performed using the gpu by simply converting the data from an array to a gpu array using the gpuArray() function. Once this is performed, the data must then be sent back to the regular memory before use in some functions by the gather() command.

Image processing times can be accelerated by between 10 and 100 times depending on the image size and the type of image processing function performed. In this example, linear filtering a 41 slice image can be reduced from $30\,\mathrm{s}$ to approximately $0.3\,\mathrm{s}$.

## 3.4 Cell segmentation

Detection of cellular boundaries can be accomplished via region of interest processing in the Matlab Image Processing Toolbox. In cases where the cellular boundary is low contrast this becomes the only way to segment cells. This method requires the user to manually draw the cell boundary around each cell of interest which could number in the hundreds or thousands. This is the main bottle neck of time in analysis.

To overcome this a number of automatic segmentation methods were attmepted. In general, automatic detection is easiest when there is high contrast at the boundary of the cell. The downside to using a non-permeable method of labeling such as FISH is that it requires the digestion of the cell wall. THis lowers the contrast and makes automatic detection more difficult. Another challenge for automatic detection is separating close cells or mother cells from daughter buds. Since higher throughput is desired it is necessary to take images where the field of view is crowded with cells numbering between 50 to 100. In this case the boundary may have high contrast but the contrast at the contact point between the cells is lower.

### 3.4.1 Sobel edge detection

The Sobel filter is a popular form of edge detection and edges are found using gradient filtering. This and several other filters are built into the Matlab image processing toolbox. The choice of the Sobel filter is done based on the qualitative performance relative to the other options such as the Canny and Prewitt filters. In particular the goal was to detect cell boundaries, and it was therefore more important that the edge of the cell be detected than any organelles or other shapes in the DIC image. The Sobel filter seems to provide the least amount of extra edges to remove during the process of cell segmentation. The Prewitt filter performed similarly to the Sobel filter but needed more gaps connected and the Canny filter produced a lot of edges internal to the cell. The Sobel filter consists of two kernals that are

convolved with the image with one being the rotation of the other.

$$G_x = G_y^T \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \tag{3.6}$$

The Sobel filter primarily detects the steepest gradients in the image which are at the location of the cell wall and membrane. This filter detects cells quite well when the cells have high contrast at the cell wall. The raw images are smoothed using a Gaussian filter with a narrow width chosen to blur small areas where the cell boundary might appear disconnected and then edge detection is performed (Figure 3.17A).

After edge detection, small features are removed to improve the performance of subsequent binary morphological operations on these edges (Figure 3.17B). The object size filtered image is then dilated using a structuring element that will expand the width of the edge enough to close small gaps without merging the cell boundaries (Figure 3.17C). A good choice for this parameter seems to be approximately five pixels with about 50 cells per image.

Objects that are connected are then filled (Figure 3.17D) and this image is further filtered by object size and shape to exclude detected ojects that are too large or non-elliptical (Figure 3.17E).

When cells are close together or boundaries are low contrast the method of detection under performs substantially.

### 3.4.2 Watershed segmentation

One form of cell segmentation is based on the convolution of the DIC image with a reference cell. Assuming the population is of fairly uniform morphology then the convolution produces strong peaks when the reference cell overlaps with Watershed segmentation is a procedure to use a feature of the image to seed an intensity distribution. This can be itera-

tively assigned to a basin in this distribution. The choice of basin is based on the gradient of the intensity. Essentially, each seed pixel is assigned a label. If the neighbors to the seed pixel only have neighbors with the same label or no label then the neighboring pixels are assigned to the label. This process is repeated until it is no longer possible to make assignments. The pixels with no label are then the edges of the basins.

The DIC image does not have a gradient suitable for this process. Also, it is difficult to make a reference cell that can be convolved with the image to produce an intensity distribution suitable for watershed segmentation. Therefore, the DAPI channel is used to seed an image of the black and white distribution of pixels (Figure 3.18). The major flaw with this approach is that the size and location of the nucleus does not directly correspond to the shape and location of the cell since the nucleus is offset from the center of the cell in most cases. The watershed algorithm also tends to oversegment the image (Figure 3.18C).

### 3.4.3 Edge detection with gap closure

Edge detection with a Sobel filter can be improved by locating the ends of the detected edges and joining them without using image dilation. In this case each edge has the end-points located and a circle of some radius is placed at the ends and if they overlap the ends are connected along the shortest path (Figure 3.19B).

These holes are then filled (Figure 3.19C) and then the edges are subtracted from the filled image to separate the filled regions (Figure 3.19D). In some operations Matlab considers connectivity to be adequate within the 8-pixel boundary around a pixel so that the edges are then shrunk by 1 pixel by erosion using a disk of radius 1 as a structuring element (Figure 3.19E).

This method performs much more successfully than any previous attempt at automatic segmentation, but was only recently implemented and as such did not see any use in the data presented in the following chapters.

## 3.5 Detecting spots in cells

### 3.5.1 Detecting peaks that are sparse

The simplest way to detect spots that are sparse is to apply an intensity threshold and then count the spots that are represented as connected components in the binary image. The most important features for this type of detection are good signal to noise and low spot count. In yeast cells, the size of the cell makes it difficult to perform this type of spot counting accurately when the number of spots in the cell exceeds between twenty to thirty. Additionally, this method tends to under count because any spots that are close end up becoming connected in the thresholded image. This also creates a balancing act when choosing a threshold. On the one hand, a high threshold will separate the most intense spots that are near each other. Whereas, a lower threshold will cause those spots to group, but will be able to detect spots that are poorly excited. This makes a uniform excitation very important since that has a direct impact on the ease of choosing a appropriate threshold.

To improve the detection of spots via this method it is also recommended to perform some type of image enhancement, which at a minimum is the application of a LoG filter to the data before applying a threshold.

The intensity of the spot detected by this method is then the sum of the pixels in the connected region.

### 3.5.2 Detecting peaks with Gaussian fitting

To perform peak detection with fitting, the image is first searched for local maxima. To improve speed, the image can be masked to search only the most intense regions of the image by applying a threshold on the LoG filtered version of the image.

This method is different from the first since each connected region of intensity is searched for local maxima rather than being treated as a single spot. A further difference is that the intensity threshold chosen to select the indices to search does not need to

be as strict as the one used to count spots since the goal is to accelerate the search for local maxima.

One approach that was tried was to then take the detected local maxima and extract the intensity of the spot based on subtracting the mean of the annular region around the spot from the sum of the pixels around the peak that were weighted based on a fixed Gaussian profile. This method is reasonably quick and effective, but yields poor intensity information when spots are close due to the neighboring spot overlapping with the annular region used to approximate the background intensity.

Once each maxima is found it is then fit using a nonlinear least squares fitting function in Matlab and a 3-dimensional Gaussian profile. The coordinates, amplitude and $\sigma$ are extracted as the relevant parameters of the spot.

A version of this method is used for most of the data in the following chapters.

### 3.5.3 Detecting peaks via iterative removal of spots

The CLEAN algorithm is used to detect FISH spots. In this case, the maximum intensity peak within the cell boundary is selected and a Gaussian profile is subtracted from the location, which represents the intensity profile of a single fluorophore scaled to the local intensity of the cell. The damping parameter and stopping point for this algorithm are selected per strain with the goal being to obtain a quality reconstruction with a minimum of residuals (Figure 3.23. These parameters determine the amount of intensity subtracted from the 'dirty' image. The damping parameter is selected by choosing a number that separates the fluorophores intensity profile from the background. The stopping parameter is selected by choosing an arbitrary cutoff which should represent the minimum intensity of an accepted spot. In the case of variability of intensity across the image, the damping parameter should be chosen so that any single spot is subtracted leaving the background below the intensity cutoff.

This method seems to be the best for spots that are separated by a distance near the

diffraction limit. This is particularly the case in the chapter regarding TLC1 where the transcript in question localizes to the nuclear membrane.

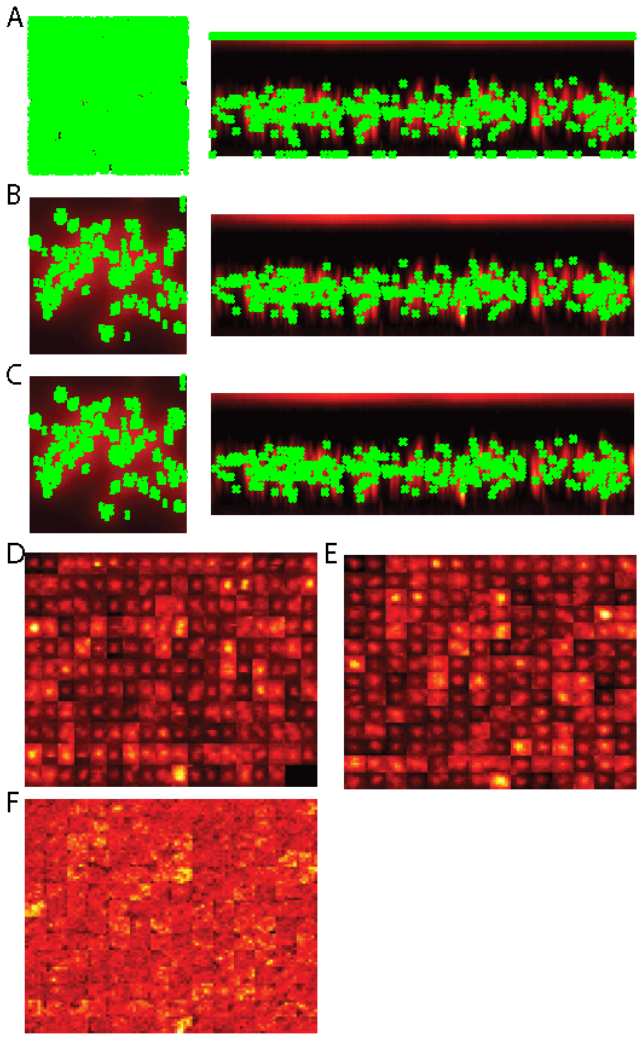


Figure 3.3: Fish-quant parameter determination. A. All spots. An XY and XZ projection of the image with all detected spots is shown. Fish-detects many spots at the coverglass with the top z-slice being the coverglass side of the chamber. B. Spots filtered in Z. The spots outside the Z-threshold are removed. C. All spots except for transcriptions sites and nuclear spots. All spots appearing in the segmentation of the nucleus and selection of the transcription sites are removed. D. XY projection of spots. The maximum intensity region around the spot center is shown in an image constructed from the XY projections of all of the spots. E. XZ projection. The orthogonal view is constructed for each spot and shown as an XZ projection of all spots in a single image. F. Residuals. Since the spot represents the fitted and background subtracted spot the residuals are shown for reference.

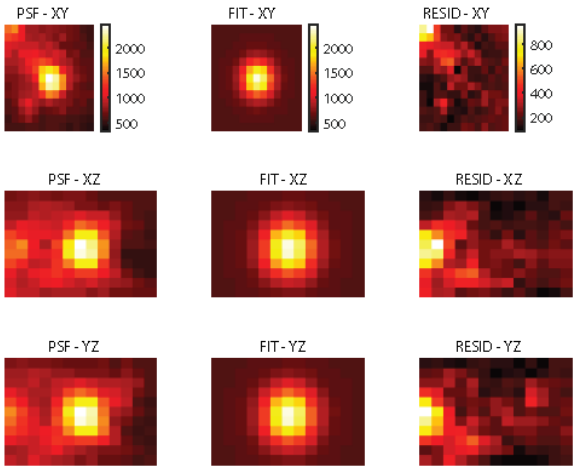


Figure 3.4: Fishquant PSF reconstruction. The detected spot in the smoothed and background subtracted raw data is shown on the left. The PSF is modeled as a 3-dimensional gaussian and is shown as the center column. The residuals of the PSF fit are shown as the left column.

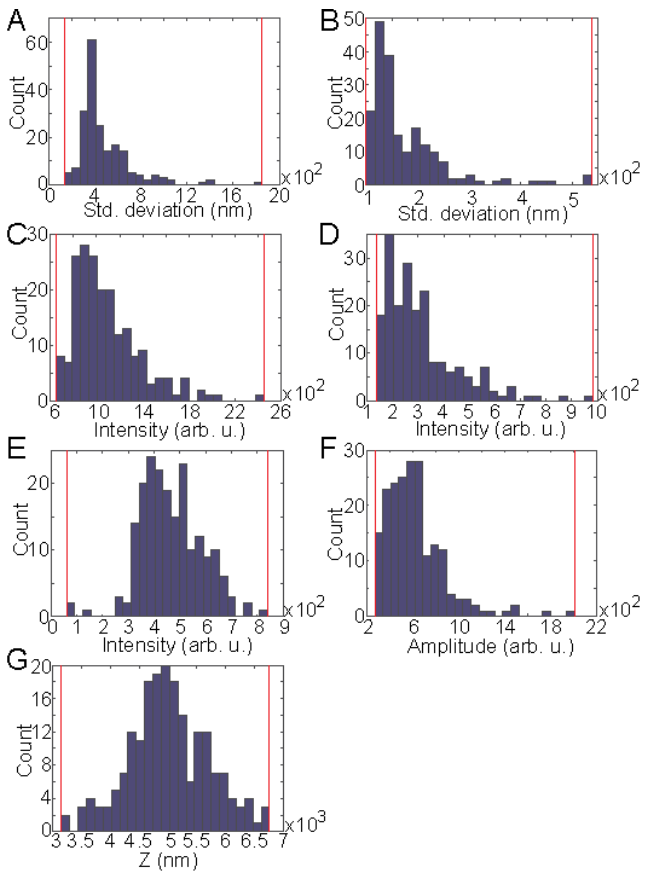


Figure 3.5: Arbitrary parameter selection. A. Standard deviation of spot locatoin in XY. The maximum and minimum (red) acceptable widths of the Gaussian fit in XY are selected. B. Standard deviation of spot location in Z. The max and min of the Gaussian profile in Z are selected. C. Intensity threshold. A second intensity threshold is selected which applies to the raw intensity. D. Intensity threshold using filtered image. Another set of parameters allows the max and min of the intensity after filtering with a 2-dimensional Gaussian filter. E. Background intensity threshold. The intensity of the Gaussian approximation of background intensity is used as yet another set of parameters. F. The amplitude of the Gaussian fit. The amplitude determined by fitting is used to threshold spots. G. Z position. The location of the detected spots in the Z slices is used to threshold spots.

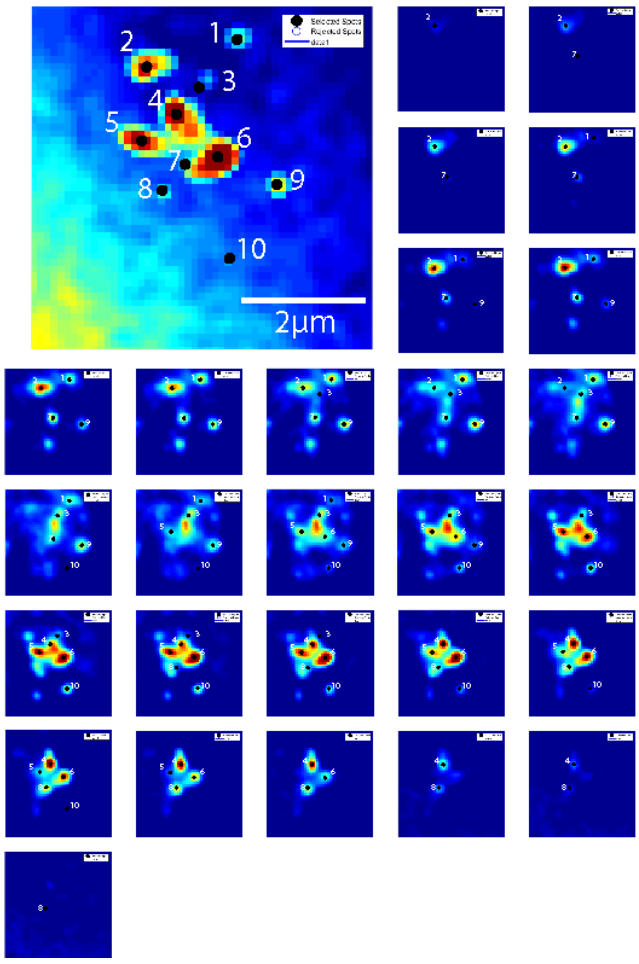


Figure 3.6: Example slices of a single cell spot detection. The maximum projection of the z-stack image is shown with spots marked and labeled with numbers to identify them in the individual slices (labeling is not a feature of Fish-quant). The individual slices are shown to illustrate the poor spot identification for cases where spots are clustered but not localized to the the same sub-voxel volume. It is clear in comparison to the PSF of the microscope and detection of individual probes shown previously that many of these intense regions are due to spots that are near the Abbe limit making separate detection of peaks very challenging. It is also unclear why the algorithm fails to identify some peaks that seem to meet the criteria even though the minimum separation between spots is chosen to be small (1 pixel).

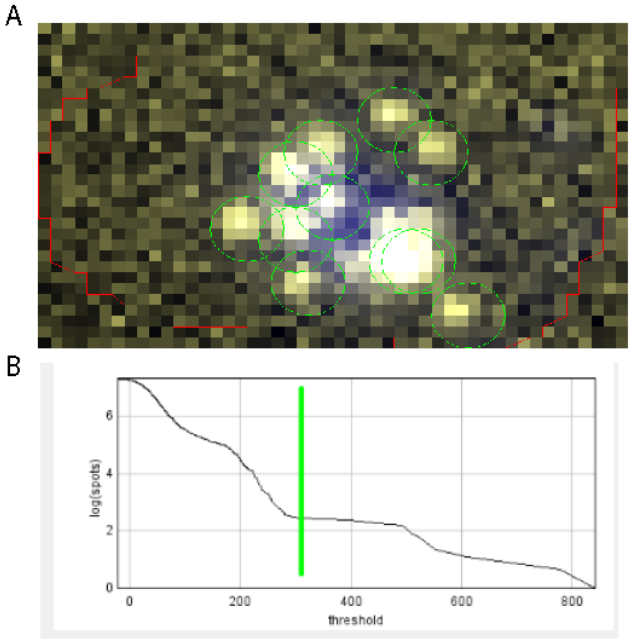


Figure 3.7: Star Search. A. Detected spots. The output of a single cell for the example data is shown. Star Search displays the maximum projection. Detected spots are circled in green and the cell boundary is outlined in red. The boundary is hand drawn and the appropriate number of spots is chosen for each segmented cell by hand. B. Thresholding. The number of spots detected is shown as a function of the chosen threshold and is used as the means of choosing the correct threshold. This analysis is performed on the maximum projection.

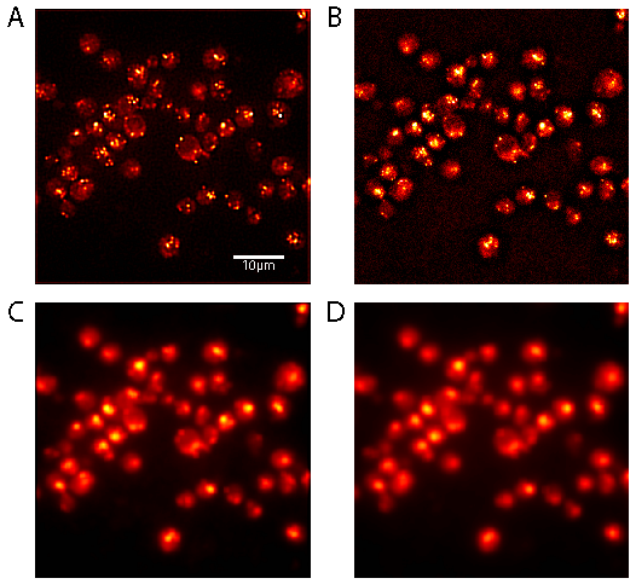


Figure 3.8: Background estimation. A. Background subtraction via guassian filter. A wide gaussian filter ($\sigma = 15$) is applied and the result is subtracted from the raw data that is blurred by a narrow gaussian ($\sigma = 1.5$). B. Weighted background subtraction. The wide filtered image and narrow filtered image are both multiplied by coefficients whose sum is 1. The difference is then taken and shown. C. Wide guassian filter. Background is estimated by blurring the image with a wide filter. D. Averaging filter. Another type of way to estimate the background is an averaging filter which replaces the center pixel with the average of all of the pixels in the chosen window. The kernal of both the gaussian and averaging filter can be seen as artifacts in the respective images.

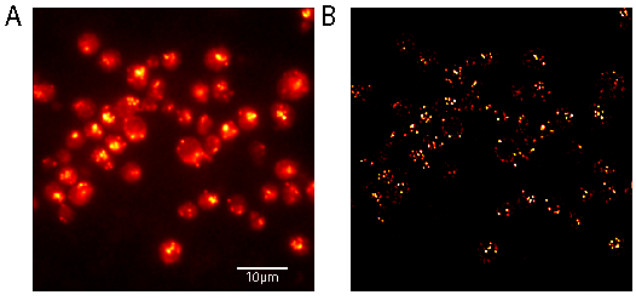


Figure 3.9: Laplacian of Gaussian. A. Raw data. The center slice of the Cy3 image is shown as a reference. B. LoG filter. The LoG filter enhances the center pixel value and reduces the pixels around the center making it ideal for detecting edges and enhancing punctate spots.

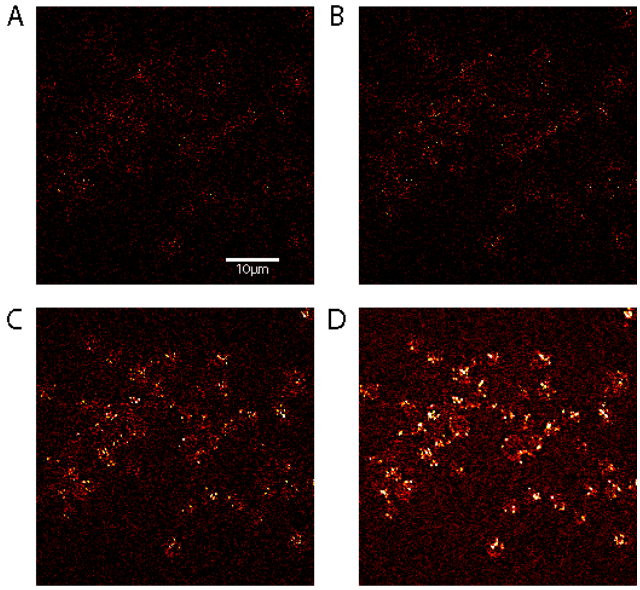


Figure 3.10: Top-hat transformations of different sizes. A. 2x2 structuring element. In this type of operation, variations in the intensity that are smaller than then the structuring element are preserved. B. 3x3 diamond structuring element. As the size of the structuring element is increased or the shape changed more intensity is preserved. C. 4x4 diamond structuring element. A larger size makes spot locations more obvious. D. 11x11 diamond structuring element. When the size of the structuring element is larger than the objects of interest the background intensity becomes more apparent.

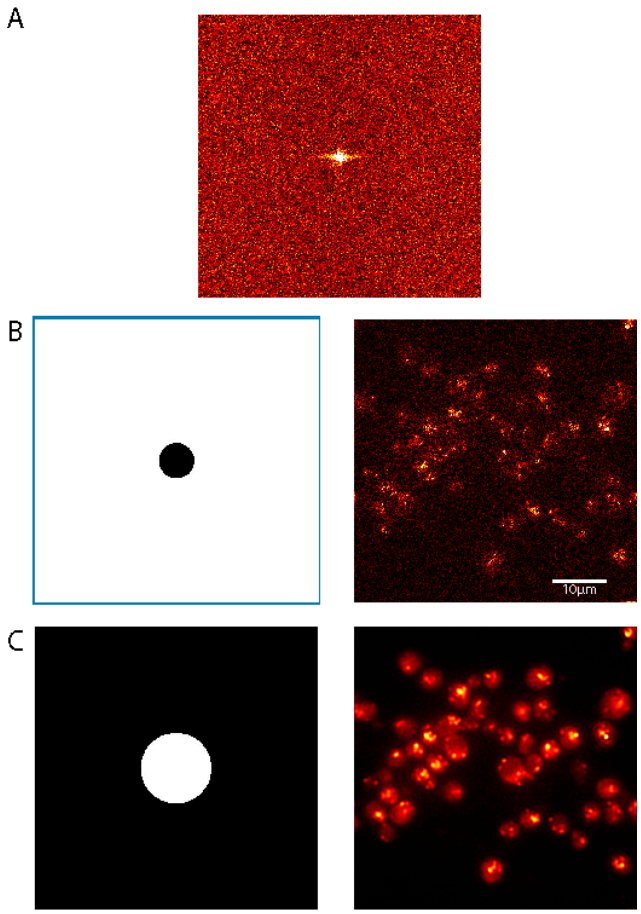


Figure 3.11: Filtering in Fourier space. A. The fast Fourier Transform. The image is converted to Fourier space by the fft algorithm. The resultant image is shifted to be centered. B. High-pass filter. High frequency variation in the image can be passed by creating a logical image that sets the values near the center to zero. Noise and spots can be seen in the result of the inverse fft. C. Low-pass filter. A low-pass filter sets all values above a cutoff to zero. The inverse fft shows that the information returned represents the background intensity of the image.

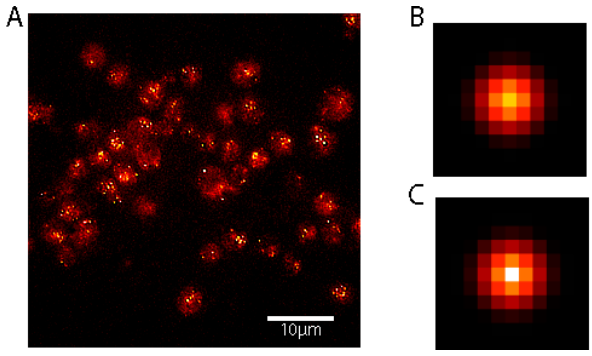


Figure 3.12: Blind deconvolution. A. Result of blind deconvolution. In this algorithm the PSF and image are both reconstructed. B. PSF input. The input PSF is assumed to be Gaussian. C. Reconstructed PSF. The input PSF is restored iteratively yielding the image shown.

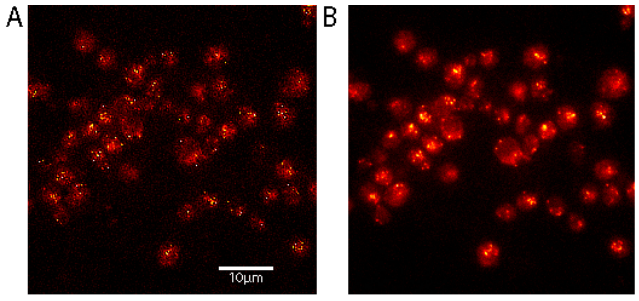


Figure 3.13: Lucy-Richardson deconvolution. A. Result of Lucy-Richardson deconvolution. This algorithm was applied and example slice of data is shown. B. Raw slice. The raw data is shown for reference.

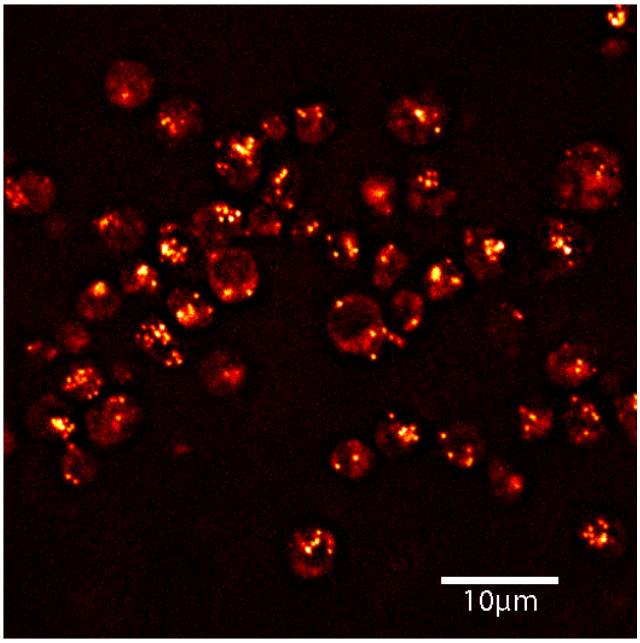


Figure 3.14: Nearest-neighbor deconvolution. Background intensity in the focal plane is assumed to be due to blur of the neighboring focal planes. The images one step above and below are blurred using a Gaussian filter and then subtracted from the central slice.

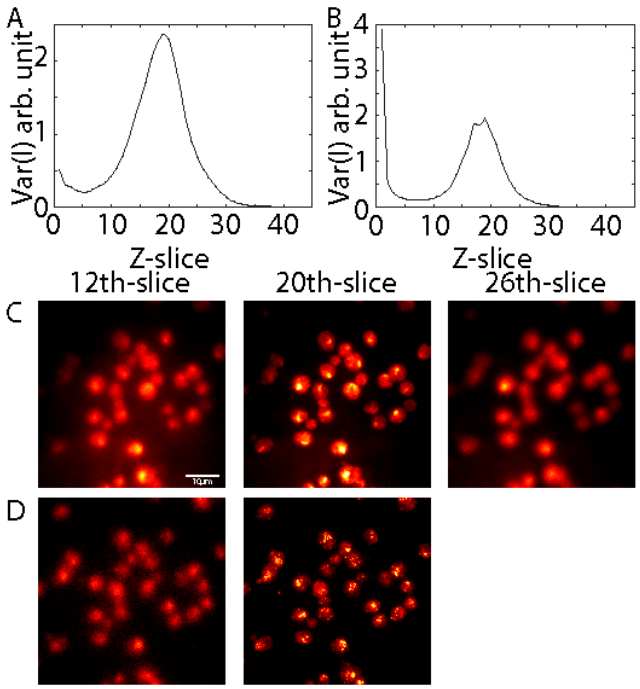


Figure 3.15: Variance of image intensity. A. Variance in Nucleolar label ITS1. The variance is calculated for each slice of a z-stack image of Cy5 labeled RNA that target a ribosomal RNA intron ITS1. This value increases as the image becomes in focus. B. Variance in TLC1 RNA signal. The variance of Cy3 targeting the non-coding RNA TLC1 is calculated. This also peaks when the image is most in focus. C. Center and boundary images of ITS1. When the variance is low the z-slice is clearly out of focus (right, left) and the nucleoli can be observed in focus when the variance is high (center). D. Center and boundary images of TLC1. The variance in an image for a small number of spots also is maximum when in focus (center)

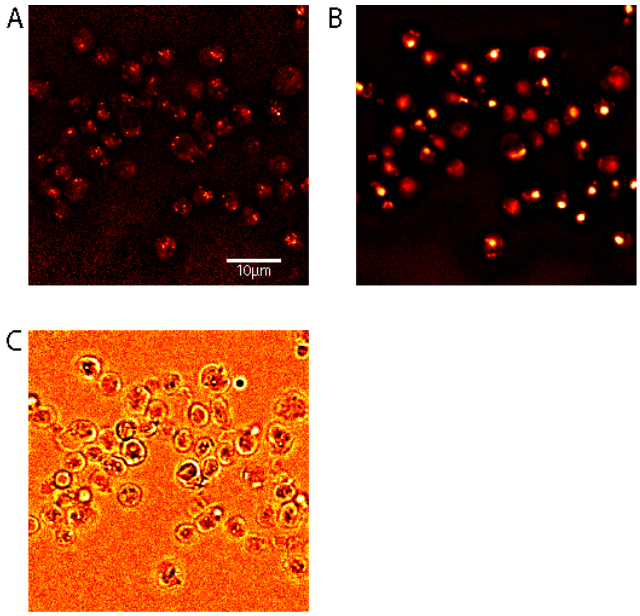


Figure 3.16: Example images processed by CIDRE. A. mRNA after CIDRE. This algorithm flattens the image by energy minimization yielding good contrast and signal to noise for mRNA spots. B. DAPI after CIDRE. The image processing flattens the DAPI channel which was fairly uniform already due to the broad illumination profile of the $375\,\mathrm{nm}$ LED. C. DIC image after CIDRE. The DIC image is significantly improved by CIDRE processing and probably shows the most obvious improvement due to the non uniformity of the DIC illumination.

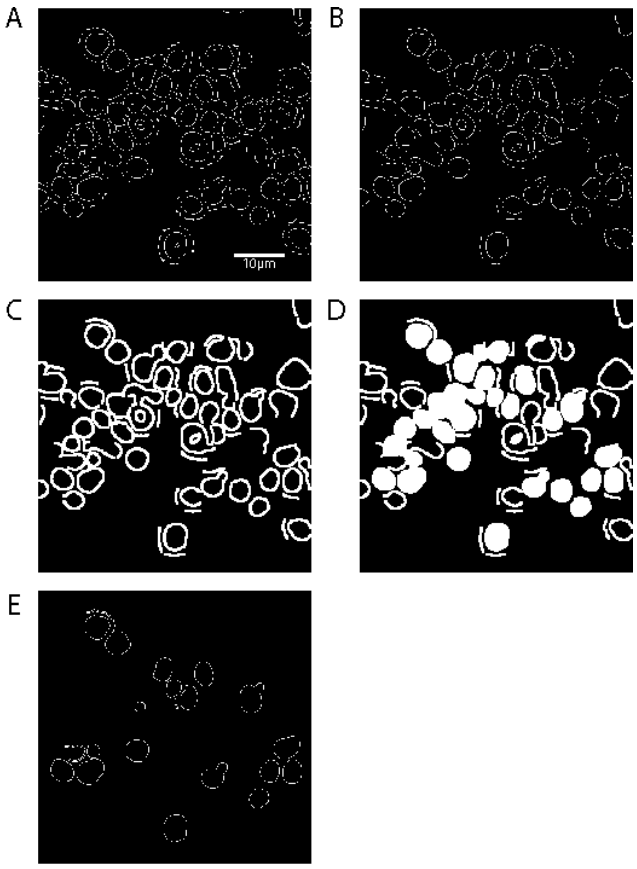


Figure 3.17: Edge detection via Sobel filter. A. Sobel filter result. The edges in the image are shown as a binary image. B. Removal of small features. Small connected objects in the image are removed to make successive binary morphological operations more successful in detection. C. Image dilation. The edges are dilated using lines designed to close gaps in the image. This dilation is limited by the ability to uniquely separate the objects in the image. D. Hole filling. Holes are filled in the dilated image and then the image is succesively opened and closed to remove unfilled edges. E. Resultant borders. The borders of the filled objects that are below a size threshold are shown.

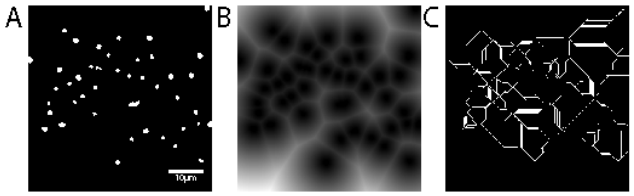


Figure 3.18: Watershed segmentation. A. Nucleus location. The DAPI channel is used as a reference for the presence of a cell. This image is converted to a logical image using an intensity threshold. B. Binary distribution of the image. The image is converted to a distribution of weights using the bwdist() function. This image is complemented to created a suitable image for watershed processing. C. Watershed edges. The result of watershed segmentation shows that the DAPI channel is a poor choice of seed for this type of segmentation.

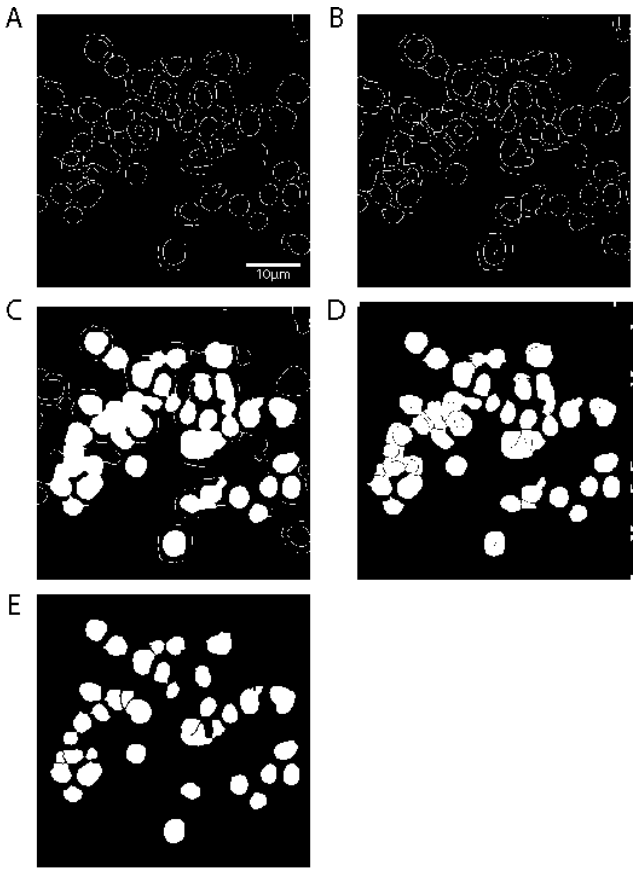


Figure 3.19: Gap filling edge detection. A. Sobel edge detection. The cleaned edges in the DIC image are shown as a reference for the starting input. B. Gap filling. Each connected object is joined by calculating its end points and measuring the distance between them. Gaps that are ten pixels or smaller are joined. C. Hole filling. Holes are filled using binary morphological operations. D. Edge subtraction. The difference between the filled image and the edge image is show. E. Shape filtration and erosion. The image is dilated and eroded and then each object has its eccentricity measured. Objects that are less round are removed.

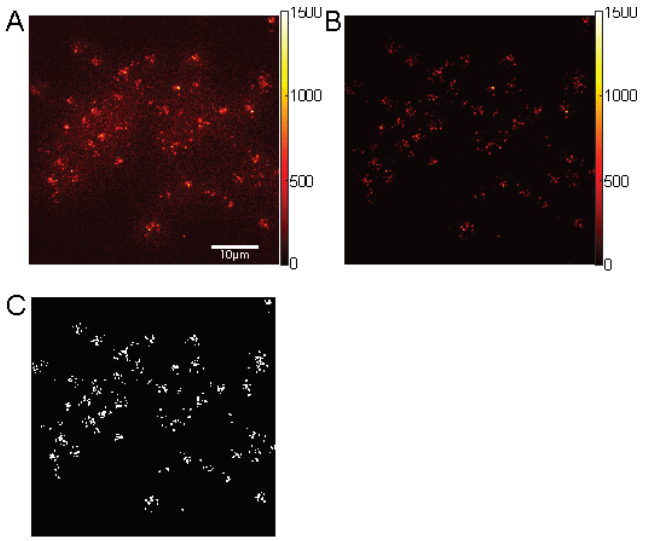


Figure 3.20: Detecting spots using intensity thresholding. A. The maximum projection of the raw data is shown for reference. B. Laplacian of Gaussian filtration. A LoG filter is applied to the image and the result is shown. The width of the filter is deteremined by the fitted size of a spot. C. Intensity thresholding. The filtered image is converted to a logical image using an intensity threshold chosen to split peaks from background. When spots are sparsely distributed they are easily identified, but this method is poorly sensitive to clustered spots.

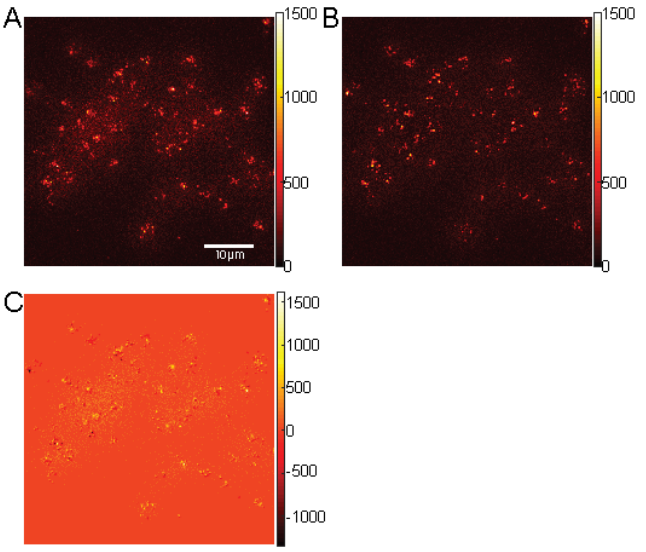


Figure 3.21: Peak detection with spot fitting. A. The maximum projection of the raw data is shown for reference. B. Reconstruction with fit parameters. The image is reconstructed by taking the background and calculating intensity where spots were detected using the fit parameters of the gaussian fit to each spot. C. Residuals of reconstructed image. The residuals of the difference between the two projections is shown as a reference.

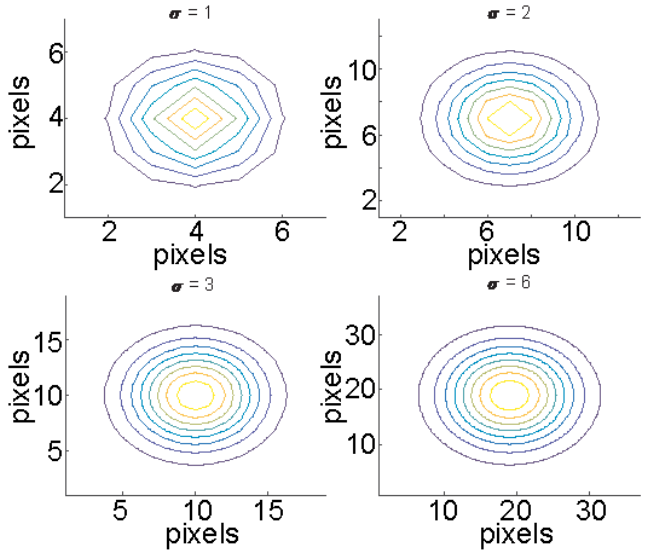


Figure 3.22: Quality of Gaussian approximation. When fitting a spot with a Gaussian profile, the standard deviation of the spot intensity is used to determine the width of the Gaussian profile. For spots with a narrow width in pixels the data is very discontinuous and the fit is less accurate. As the width increases the quality of the fit improves since the data because smooth. One approach to improve this fit is to interpolate between the data points.

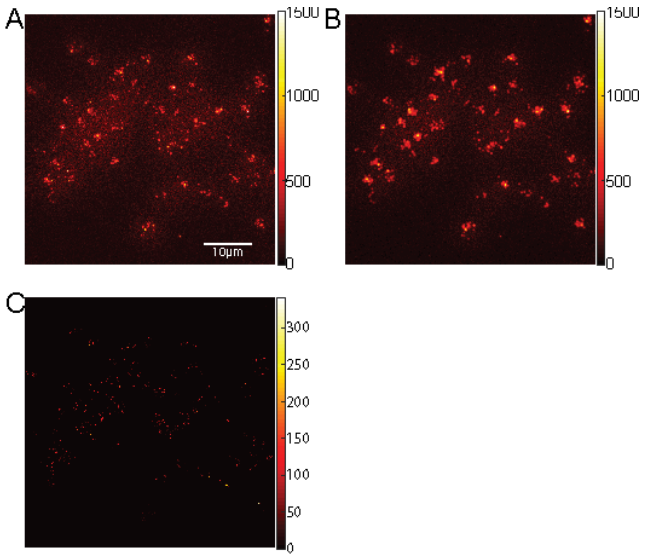


Figure 3.23: CLEAN. A. The maximum projection of the raw data is shown for reference. B. CLEAN result. The projection of the reconstructed image from the combination of detected spots and the 'clean' image of the background is shown. C.Residuals from CLEAN. The residuals of the difference of the two projections are shown as a reference for the reconstruction.

# CHAPTER 4
# SINGLE FLUOROPHORE FISH

## 4.1 The protocol

The following protocol has been adapted from Wadsworth, et al[59] in three variations; a single probe method and a method using FRET that both appear on the online resource Bio-protocol.org, and a cost effective method that will appear in the Methods in Molecular Biology series. The following is taken from the single probe protocol.

### 4.1.1 Day 1

At the end of the day, inoculate yeast cells into 50 ml of liquid media in a culture flask from cells actively growing on a plate.

### 4.1.2 Day 2Cell fixation and permeabilization

1. Measure Cell OD using a spectrophotometer at $OD_{600}$ by placing $1\,\mathrm{mL}$ of cell culture in a cuvet.

2. Once cell $OD_{600}$ has reached 0.6, decant cells into a $50\,\mathrm{mL}$ Falcon tube and pelleted by centrifuging at 671 x g for $5\,\mathrm{min}$ and aspirated.

3. Resuspend the pellet in $10\,\mathrm{mL}$ of ice cold ($4\,^{\circ}\mathrm{C}$) methanol for $10\,\mathrm{min}$ for fixation.

4. Cells are pelleted and resuspended in ice cold Buffer B twice and aspirated.

5. Resuspend the cells in $1\,\mathrm{mL}$ of Spheroplasting buffer and transfer to a $1.75\,\mathrm{mL}$ microcentrifuge tube and add $2\,\mu\mathrm{L}$ of $5\,\mathrm{units}/\mu\mathrm{L}$ of zymolyase and gently pipette to mix.

6. Incubate the cells for 30 min or until the $OD_{600}$ of $100\,\mu L$ of cells added to $900\,\mu L$ of deionized water shows a reduction of 30% from the initial OD after $1\,min$, which demonstrates cell lysis.

7. Pellet the cells and aspirate. Centrifuge at no more than 268 x g since they are fragile once they are spheroplasts.

8. Wash the cells two more times in ice cold Buffer B and aspirate (wash means spin down to a pellet, aspirate liquid and resuspend.)

9. Resuspend the cells in $1\,mL$ of 70% ethanol and keep at $4\,°C$ for a minimum of $1\,h$ to overnight.

#### 4.1.3 Day 2 cont./Day 3hybridization

1. Pellet cells at 268 x g and wash twice each with $1\,mL$ Wash Buffer and aspirate. Note: Wash buffer should be prepared fresh and formamide should be warmed to room temperature before opening.

2. Dilute the probes to $1\,\mu M$ in $10\,mM$ Tris-HCl, pH8.

3. Prepare a mixture of hybridization buffer and probes based on the working concentration determined by titration.

4. Resuspend cells to a final volume of $100\,\mu L$ in the probe-hybridization buffer mixture by gentle pipetting.

5. Then wrap samples with aluminum foil and place in the incubator at $30\,°C$ overnight.

#### 4.1.4 Day 3/Day4Slide Preparation

1. Prepare Imaging Buffer immediately before use.

2. Wipe the slides with ethanol; or (optional) clean slides and coverslips in a plasma cleaner for 10 min. Note: The slide should be clean of dust and other particles by wiping with ethanol. Any air bubbles will severely impact the performance of the Imaging Buffer.

3. Mix $2.5\,\mu L$ of Imaging Buffer with $2.5\,\mu L$ of cells and place on the coverslip.

4. Place the coverslip on a slide and seal with epoxy along the edges. Note: Slides should be kept in a dark place while not on the microscope. Several slides can be prepared simultaneously. Once sealed, the performance of the imaging buffer will not degrade for several hours

## 4.2 sFISH

The following is drawn from work published in Wadsworth, et al[59].

### 4.2.1 Strain construction

We constructed multiple PHO5 promoter variants of budding yeast following the protocol used in our previous study[60]. These variants share a high affinity Pho4 binding site in the exposed region between nucleosome -2 and nucleosome -3, thus belonging to a family of HX promoter variants[61]. These variants differ in their DNA sequence in the nucleosome -2 region (Supplementary Figure 1 and Supplementary Table 1). The PHO5 open reading frame was then replaced with a yellow fluorescent protein (yEVenus) gene by homologous recombination. To achieve constitutive expression of yEVenus, the PHO5 promoter variants were mated with the pho80$\Delta$ strain[62]. yEVenus levels of these promoter variants were quantified using the epi-fluorescence microscope (Supplementary Figure 1B). Yeast strains with different ploidies (2N, 3N, 4N) were from Dr. David Pellman.

### 4.2.2 Sample preparation

Our procedure closely follows the protocols by Youk et al.[63] and Raj et al.[52] with some modifications. Yeast cells are grown overnight to a final OD600 of 0.5 in $50\,\mathrm{mL}$ of SD complete medium. Cells are fixed and then kept at $4\,^{\circ}\mathrm{C}$. Fixation is performed either by treating the cells with $2\,\%$ v/v formaldehyde or with methanol for 10 minutes. In the case of methanol, cells are spun down, and the pellet is resuspended in methanol. The cells are washed twice with Buffer B ($1.2\,\mathrm{M}$ sorbitol, $0.1\,\mathrm{M}$ potassium phosphate) at $4\,^{\circ}\mathrm{C}$ after fixation. Cells are then resuspended in $1\,\mathrm{mL}$ of spheroplasting buffer ($10\,\mathrm{mL}$ Buffer B, $100\,\mu\mathrm{L}$ $200\,\mathrm{mM}$ vanadyl ribonucleoside complex from New England Biolabs). $2\,\mu\mathrm{L}$ of zymolyase at $5\,\mathrm{units}/\mu\mathrm{L}$ (Zymolyase-20T at $21\,000\,\mathrm{units}/\mathrm{g}$ from Seikagaku Business Corporation) are added into the mixture of cells, and then they are incubated at $30\,^{\circ}\mathrm{C}$. The amount of digestion is determined by measuring OD600 of $1\,\mathrm{mL}$ of a suspension containing $100\,\mu\mathrm{L}$ of the cell sample until the OD600 has decreased by about $30\,\%$, which is typically 15 minutes. Cell wall digestion can be verified by allowing $10\,\mu\mathrm{L}$ cells to settle to the surface of a multi-well plate and adding $100\,\mu\mathrm{L}$ DI water to the well. Within a few minutes the majority of the cells should be lysed when inspected under a microscope. If the cells do not lyse, then the procedure has not effectively removed the cell wall. Following this treatment the cells are washed twice in Buffer B at $4\,^{\circ}\mathrm{C}$ and stored in $70\,\%$ ethanol for at least one hour.

Hybridization is performed by washing the cells twice with $1\,\mathrm{mL}$ of wash buffer containing $10\,\%$ formamide and $330\,\mathrm{mM}$ salt as SSC buffer. This step loosens the protein-RNA interactions to increase RNA accessibility for probe hybridization, and decreases nonspecific probe binding[64]. Hybridization buffer is prepared in $10\,\mathrm{mL}$ volumes containing $1\,\mathrm{mL}$ 20X SSC (Ambion), $1\,\mathrm{mL}$ Formamide (Ambion), $100\,\mu\mathrm{L}$ $200\,\mathrm{mM}$ vanadyl ribonucleoside complex (New England Biolabs), $1\,\mathrm{g}$ dextran sulfate sodium salt (Sigma, D8906), $10\,\mathrm{mg}$ *Escherichia coli* tRNA (Sigma, R1753), $40\,\mu\mathrm{L}$ of $5\,\mathrm{mg}/\mathrm{mL}$ BSA (Ambion), and $8\,\mathrm{mL}$ deionized water. Cells are resuspended in hybridization buffer and the appropriate

amount of probes is added to bring the working concentration to $65\,\mathrm{nM}$ and the final volume to $100\,\mu\mathrm{L}$. Cells are then incubated at $30\,^{\circ}\mathrm{C}$ overnight. The following morning, cells are washed twice in wash buffer and left as a pellet. Slides are prepared by mixing $2\,\mu\mathrm{L}$ of concentrated cells with $2\,\mu\mathrm{L}$ of imaging buffer that has the same concentration of salt as the wash and hybridization buffers. $1.5$ to $2\,\mu\mathrm{L}$ of cell suspension is placed between a microscope slide ($1'' \times 3''$) and a coverslip (#1.5, $18\,\mathrm{mm} \times 18\,\mathrm{mm}$), and the chamber is gently pressed to form a monolayer of cells. The edges of the chamber are sealed with fast curing epoxy. The height of the chamber is estimated to be about $3\,\mu\mathrm{m}$. Cells slightly squeezed in the chamber remain confined during the observation period. The imaging buffer contains the PCA/PCD oxygen scavenging system[65] to extend the photobleaching lifetime of the fluorophores. The imaging buffer is composed of $10\,\mathrm{mM}$ Tris at pH8, $5\,\mu\mathrm{L}$ 20X SSC, $2.5\,\mathrm{mM}$ protocatechuic acid (PCA), $10\,\mathrm{nM}$ protocatechuate-3,4-dioxygenase (PCD), and $1\,\mathrm{mM}$ Trolox (6-hydroxy-2,5,7,8-tetramethylchroman-2-carboxylic acid).

### 4.2.3 Observation of single fluorophores in vivo

While the current FISH approaches have been successfully applied to quantify mRNA level at the single cell level, they are not applicable to studies that investigate short RNA molecules or changes in short regions of RNA because of the multiple-probe requirement. To circumvent these limitations, we sought to develop a single-probe FISH technique for budding yeast. The three parameters we considered as we began our modified FISH method was the dye molecule, the fixative, and imaging technique. We chose Cy5 over Cy3 because yeast autofluorescence is lower in the red channel than the green [66]. In addition, we tried two different fixatives that may differentially contribute to the signal-to-noise of Cy5 [67].

We first acquired fluorescence images of cells of a positive control strain under epi-illumination. We were able to observe isolated fluorescent spots, but the signal-to-background ratio was poor. Hence, we tried illuminating the cells using an inclined geometry (Fig-

ure 2.4B). Highly inclined illumination excites a thin slice inside the cell, which leads to higher excitation intensity and lower out-of-focal plane background compared to epi-illumination[37]. Using this setup, we observed a significant enhancement in the signal-to-noise ratio (Figure 4.1B-C). The oxygen scavenging system was also critical for detection of the isolated fluorescent spots. Two observations indicate that most of these spots arise from a single Cy5 molecule. First, their fluorescence intensities are comparable to the fluorescence intensity of single Cy5 molecules nonspecifically bound to the surface. Second, upon continuous excitation, most spots disappear in single photobleaching steps, consistent with a single Cy5 molecule (Figure 4.14).

### 4.2.4 Correlation between spot count and mRNA level

To test the linearity of our sFISH protocol, we performed FISH on four different strains (Figure 4.2A) that express yEVenus at four different levels (Supplementary Figure 1B). We did not use a promoter inducible by external factors as different induction conditions may differentially affect the hybridization efficiency.

We compared the measured spot counts with yEVenus levels (Figure 4.2B). The spot count ranged from 2 to 32. We observed a good correlation between spot count and protein level, which argues that fluorescent spots are generated from specific hybridization of the probe to the target mRNA. It also indicates that our spot counting algorithm works well in the range of transcript levels tested.

As another control, we performed sFISH on chromosome copy number variants (1n, 2n, 3n, and 4n). We chose to probe a constitutively expressed gene KAP104 (Supplementary Tables 4, 5), which has been used as a reference gene in other studies[28, 68]. As expected, the number of spots monotonically increased with the ploidy (Figure 4.2C), but interestingly, the relationship was not linear, that is, doubling the ploidy did not lead to doubling of the number of spots. This apparent sub-linear relationship could be due to the loss of extra chromosomes or some compensation effect, which will be the subject of future

investigation.

In other controls, we were able to mask these sFISH spots with an unlabeled competitor probe in a concentration dependent manner (Supplementary Figure 4C and D), which confirms that the observed sFISH spots are due to the fluorescent probe hybridized to the KAP104 transcript. We also compared sFISH with mFISH (Supplementary Figure 4A). mFISH yielded an average of 9.4 spots per cell, similar to the number previously reported[28, 68]. In comparison, sFISH yielded less spots per cell than mFISH (6.5, Supplementary Figure 4B, which suggests that sFISH can detect the KAP104 transcripts with detection efficiency ($p$) of about 64% taking into account the false positive rate (0.5 spots per cell, Supplementary Figure 4D).

### 4.2.5 Methanol vs. formaldehyde

We initially tried methanol as a fixative following a fast FISH protocol[69], and noticed that fluorescence images looked more clear than when using formaldehyde as the fixative. Hence, we performed a systematic comparison of the two fixatives. We first compared the number of background spots in a negative control strain lacking the yEVenus gene between the two fixatives (Supplementary Figure 8). We found on average 3 spots per cell in formaldehyde-fixed cells compared to 0.3 spots per cell in methanol fixed cells, which indicate that formaldehyde fixation causes more nonspecifically bound or trapped Cy5 probes inside the cell (Figure 4.4A). Next, we compared the number of spots in a positive control strain with the lowest expression level of yEVenus. In this strain, we detected on average 10 spots per cell with methanol and 3 spots per cell with formaldehyde (Supplementary Figure 7). This result suggests that methanol fixation not only reduces nonspecific binding but also increases the rate of specific binding to the target mRNA.

With methanol-fixed cells, we not only detected more spots per cell in the positive control strain, but also detected more fluorophores per spot when multiple probes were used. To count the number of fluorophores per spot, we acquired time series of fluorescence from

single spots in a fixed plane until they were completely photobleached. The histograms of the number of photobleaching steps are shown (Figure 4.4D). When a single probe was used, most spots photobleached in a single step in both methanol and formaldehyde-fixed cells. Spots that photobleached in two steps are likely due to close proximity of different mRNA molecules. As the number of probes was increased (1, 5, and 30), the difference in the number of photobleaching steps between formaldehyde and methanol treated samples became more noticeable. This measurement confirms that methanol fixation allows more efficient hybridization of probes to mRNA while preserving an improved signal-to-noise ratio.

This data also allows us to estimate the hybridization efficiency of probes. Assuming that probes all hybridize with the same probability $p$, the number of probes per spot can be fitted with a binomial distribution. We fitted the binomial distribution to the two sets of data taken with methanol fixation (red dotted lines, Figure 4.4D). $p$ for yEVenus is extracted to be $61\,\%$ for 5-probe FISH, and $38\,\%$ for the 30-probe mFISH (Supplementary Table 3). The variation in $p$ between the two data could be due to the difference in probe design. In 5-probe FISH, the probes were designed to have similar melting temperatures to the probe used for sFISH (shown in Supplementary Table 2), while in mFISH, the probes are designed to have the same length with no consideration of the melting temperature. Also, poor signal of Quasar 670 used in mFISH can lead to the underestimation of the number of photobleaching steps. Nonetheless, these rough estimates set the detection efficiency in the range of $\sim 40\,\%$ to $60\,\%$.

We also characterized some apparent differences in the fluorescence properties of Cy5 due to the difference in fixatives. The fluorescence signal, which is defined as the difference between the fluorescence and background levels, was similar between the two. However, the noise, which is the fluctuation of the Cy5 signal, was significantly higher in formaldehyde fixed cells. As a result, the signal-to-noise ratio was 2-fold higher in methanol-fixed cells (Figure 4.4B). In addition to having an advantage in signal to noise ratio, methanol

treated cells exhibited a longer Cy5 lifetime (Figure 4.4C).

### 4.2.6 Detection Efficiency

To ensure that our FISH protocol operates at maximum hybridization efficiency, we increased zymolyase incubation time or the probe concentration until the spot count did not increase further (Supplementary Figures 9, 10, and 11). Even under this condition, however, our single-probe protocol is expected to underestimate the actual number of mRNA transcripts due to hindered accessibility of the target region of some transcripts. We can also estimate the effective detection efficiency ($p$) for the yEVenus transcript by increasing the number of probes. The number of spots detected per cell initially increased with the number of probes, but soon plateaued at four to five probes (Supplementary Figure 6). Assuming that each probe binds the target mRNA with probability $p$, the probability of failing to detect an mRNA molecule with $n$ probes is $(1-p)^n$. Therefore, the number of detected spots ($y$) should increase with the number of probes ($x$) as $y \propto 1-(1-p)^x$. We fitted this model to the plot of the spot count per cell vs. probe number (Figure 4.5) and extracted $p$ for the yEVenus transcript to be $53\,\%$, which is consistent with the range determined in the previous analysis (Figure 4.4).

### 4.2.7 mRNA detection via FRET

In addition to various methods used above to infer the detection efficiency $p$, we also tried to determine $p$ using Förster Resonance Energy Transfer (FRET). In this approach, two DNA probes complementary to immediately adjacent regions of the mRNA are labeled with donor (Cy3) and acceptor (Cy5), respectively, so that the FRET signal would arise only when both probes bind to the same mRNA (Figure 4.1A). By comparing the number of fluorescent spots due to FRET to the number of fluorescent spots due to direct excitation, the detection efficiency can be directly determined.

sFISH was performed with donor and acceptor probes at 1:1 ratio at the same concen-

tration used for other experiments. As controls, sFISH was also performed while leaving out one of the probes. The typical sFISH images from three different combination of probes are presented in Figure 4.8. Upon 532 nm excitation, signal in the Cy5 channel was visible only when Cy3 probe is present (left, Figure 4.8), which indicates that many mRNA molecules are hybridized with both the Cy3-probe and Cy5-probe. We confirmed that this intense Cy5 signal could not have resulted from bleedthrough of Cy3 signal into the Cy5 channel (middle, Figure 4.8) or direct excitation of Cy5 by the 532nm laser (right, Figure 4.8). Upon 532-nm excitation, spots that appear in the Cy5 channel are due to FRET from the Cy3-probe to the Cy5-probe bound to the same mRNA. On the other hand, spots in the Cy3 channel arise from mRNA molecules bound with the Cy3-probe only. We can thus estimate the detection efficiency by dividing the number of Cy5 spots by the total number of both Cy3 and Cy5 spots. Using this method, the detection efficiency is determined to be $48\,\%$.

### 4.2.8 mRNA isoform detection via sFISH

Since sFISH requires only a 20-30 nt RNA target, it can be used to differentiate mRNA isoforms that are only slightly different in length or sequence, thus offering more versatility than mFISH. As a proof of principle, we chose gene RGL1 (YPL066W), which exhibits differential usage of alternative transcription sites between glucose and galactose growth media[70] (Supplementary Figure 13). As shown in the simplified schematic in Figure 4.8B, initiation normally starts upstream of the open reading frame (ORF) of RGL1 and produces a full-length transcript, but it can also start within the ORF and produce a truncated isoform. To measure the isoform profile, we designed a pair of probes (Supplementary Figure 13, Supplementary Table 6) that flank the truncation site (solid lines, Figure 4.8B) and performed sFISH with each probe on yeast cells grown in glucose and galactose media. Since the transcription levels were too high for reliable spot counting, we instead used the total fluorescence intensity integrated over the volume of the cell as a proxy for

the transcription level. In qualitative agreement with the genome-wide transcript isoform study[70], we found that the truncated isoform is significantly enriched over the full-length isoform in galactose-containing media (Figure 4.8B).

## 4.3 Unpublished single probe experiments

### 4.3.1 Anisotropy

One of the first attempts to image yEVenus mRNA involved fluorescence anisotropy. In this configuration, the CCD was divided into 3 channels representing the two polarizations of Cy5 fluorescence and a Cy3 channel. This configuration was intended to image mRNA using FRET and to examine the polarization of the Cy5 with regards to the Cy3 molecule. Fluorescence anisotropy in this optical configuration requires linearly polirized excitation, which is determined by a linear polarizing filter placed in the excitation path. When the timescale of rotation of the acceptor is longer than the timescale of fluorescence emission, the observed light should maintain the polarization of the donor. Likewise, under direct excitation, the emission of the molecule should maintain the polarization of the excitation.

This worked adequately when the probes were attached to the surface through a 3' Biotin and bovine serum albumin (BSA) on the surface. It was shown that the molecule is on average $\sim$30% polarized (Figure 4.27). This means that the intensity in the P channel is greater than that in the S channel for almost all spots. A polarization of $\pm 1$ would meant that the intensity in one of the channels was zero, whereas, a polarization of zero means that the molecule has equal intensity in both channels. This success did not translate to probes in yeast cells. Instead, the divided intensity of the fluorophores made them much harder detect considering at the time the excitation source was the $647\,\mathrm{nm}$ line of the krypton laser and was capable of producing at maximum $0.5\,\mathrm{mW}$ of power at the BFP. With the improvements in the excitation, it is possible that this concept might be worth revisiting.

### 4.3.2 Black hole quencher

Another issue with applying the conventional smFISH protocol to a single fluorophore was the non-specific spots in the negative control. To use a single fluorophore there needs to be a minimum of false positives. This lead to using a black hole quencher as a 16 nucleotide mask for the yEVenus probe. This approach is similar to the toehold-mask design used to detect single nucleotide polymorphisms[19]. The quencher that has the most spectral overlap with Cy5 is BHQ-3. The function of a black hole quencher is a resonance energy transfer that is then dissipated through thermal radiation rather than the emission of photons by the acceptor. If the probe is hybridized to its target, then the mask should be displaced and the Cy5 should fluoresce. However, these oligos must first be annealed and then purified to be confident that this is the case. BHQ-3 labeled probes are expensive and the relative concentration required to have high efficiency in quenching without purification is unfeasible. The amount of background spots in formaldehyde fixed negative control cells was substantial even with the BHQ-3 design (Figure 4.28.) With the improvements of sFISH and fixative, it is possible that this might be worth trying again.

### 4.3.3 DNA paint on yEVenus

Another method to detect transcripts *in situ* is DNA PAINT[20]. In this method, a short DNA oligo with a single fluorophore is used to detect RNA transcripts by the repeated hybridization by an essentially infinite supply of unstable complementary probes. This method is simpler than RNA FISH because it does not require a washing step. Furthermore, it can be used for multiplexing RNA detection since the probes are intentionally unstable and therefore do not accumulate in the sample. We designed a 10 nucleotide probe labeled with Cy5 to detect yEVenus transcripts. This method requires long acquisitions at a single plane to accumulate a super resolution image of transcript locations. One major issue with this method was the available code for DNA PAINT analysis did not function and development of an analysis pipeline is a substantial effort. However, a simple accumulation

of detected peaks was easy to perform. This revealed that the amount of non-specific spots detected by DNA Paint was similar per cell to sFISH (Figure 4.26A). The diffusion of these transcripts was also tracked (Figure 4.26B) in 2D. The average duration of these trajectories is extracted as well as the average step size.

### 4.3.4 End versus internal labeling

The effect of the choice of where to label the probe with Cy5 either internally in the backbone or at the 5' end was studied by comparing the results of FISH using mpFISH to either a single end labeled probe or a single internally labeled probe. There was no apparent effect on spot detection based on these labeling chemistries (Figure 4.29). Both of them undercount transcript at a similar rate as compared to mpFISH.

Table 4.1: DNA sequence in the region of nucleosome -2.

| GC% | Sequence |
|---|---|
| 20 | AAAGTTAATTTAATATGCAATTTCTAAATTAATCGATACAACCTTAACAATAAAAGCTTGAATTAGCAAAGACTAAATTTATTATTATTGTACATTTTTTCAAATAGATTGCACATTCAAAATTATCAAATTGGTTACATTATTTTA |
| 25 | ATAGTTAATTAAATAGTAAATCTTTAAATGAATAGATACAACCTTGGCACTCAAAGCTTGGAATATTACAGACTAAATTTATAATTATATTACTTGTTTTCGAAAAGATCGCAAATGTCAAATTATCAAATTGTTCAACTTACTTGG |
| 27 | AGAGTTAATTTAATATGCAATATCTAAATTAATCGATAAAACCTTTGCAATAAAAGCTTGGACTAACATAGACTAAATTTATGATTCTATTCTCTGTTTTCAAAGAGATCGCAAATGCAAAATTATCAAATTGGTCACCTTACTTGT |
| 32 | AGAATTAATTGAATAGACAATCTCTAAATTAATCGATACAACCTTGGCACTCAAAGCTTGGACTAGAACATACTAAATTTATGATTCTTGTTTCTGTTTTCGAAGAGATCGCAAATGCCAAATTATCAAATTGGTCACCTTACTTAG |
| 35 | AGAGTTAATTGAATAGGCAATCTCTAAATGAATATATACAACCTTGGCACTCAAAGCTTGGACTAGCACAGACTAAATTTATGATTCTGGTCCTTGTTTTCGAAGAAATCGCAAATGCCAAATTATCAAATTGGTCACCTTACTTGT |
| 39 | AGAGTTAATTGAATAGGCAATCTCTAAATGAATCGATACAACCTTGGCACTCAAAGCTTGGACTAGCACAGACTAAATTTATGATTCTGGTCCCTGTTTTCGAAGAGATCGCACATGCCAAATTATCAAATTGGTCACCTTACTTGG |
| 47 | AGAGGTAATTGAACAGGCGATCTCTAAATGAATCGACACCACCTTGGCACTCAAAGCTTGGGCTAGCACAGACTAAACTTGTGATTCTGGTCCCTGGTTTCGAAGAGATCGCACATGCCAAACTATCAAGTTGGTCACCTTACTCGG |
| 50 | GGAGCCAATTGAATAGGCAATCTCTGAATGAATCGGTACAACCTTGGCACTCAAAGCTTGGGCTAGCCCAGACTCCATGTATGATTCGGGTCCCTGTTGTCGACGAGATCGCACATGCCAAATTACCAAATTGGCCACCTGACTGGG |
| 54 | AGAGTGAATTGAGTAGGCAGCCTCTAAACGCATCGACACAGCCGTGGCACTCAAAGCTTGGACTAGCGCAGCCTGAACGTACGATTCTGGTCCCCGTTTTCGAAGCGATCGCGCGTGCCAAGTCATCACATTGGTCACCTTACTTGG |
| 56 | AGGGTCCATTGGGCAGGCAAGCCCGAAAGGAACCGATACAACCGTGGCACTCAAAGCTTGGACTAGCCCAGGCTACATCGATGATTCGGGTCCCTGTTGTCGAAGAGCTCGCGCATGCCAAACCATCACATTGGTCGCCTGACTTGG |

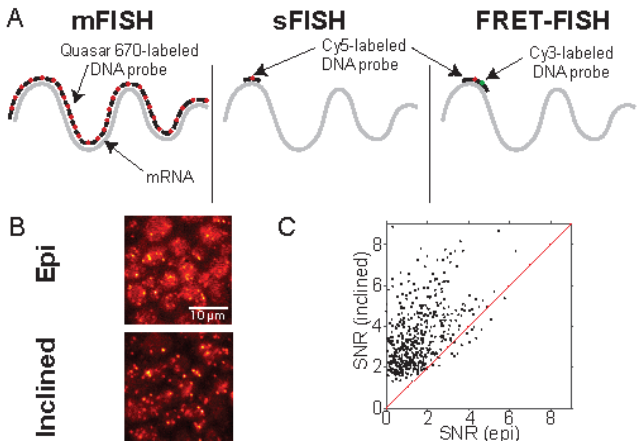


Figure 4.1: Comparison between single-probe FISH (sFISH) and multi-probe FISH (mFISH). yEVenus mRNA, which is 717-nt long, is probed by sFISH and mFISH. **(A)** Probe configurations are shown from left to right for 30 probe mFISH, sFISH and sFISH with FRET. For mFISH, we use a set of thirty Quasar 670 end-labeled probes. In sFISH, we use a single short Cy5-labeled DNA oligo probe. For FRET experiments the first Cy5-labeled probe is used in conjunction with a Cy3-labeled probe. **(B)** The images shown are of the same field of view taken with epi-illumination (top) and then subsequently with inclined illumination (bottom) using the same laser power. The bottom image taken with inclined illumination exhibits more intense spots and lower background. **(C)** Comparison of spot signal-to-noise ratio (SNR) between epi- and inclined illumination. The SNR values measured with inclined illumination is plotted against those measured with epi-illumination. Most spots are found above the red line $y = x$, which indicates inclined illumination produces higher SNR than epi-illumination. The increase in signal to noise ratio is a factor of 2.15 on average.

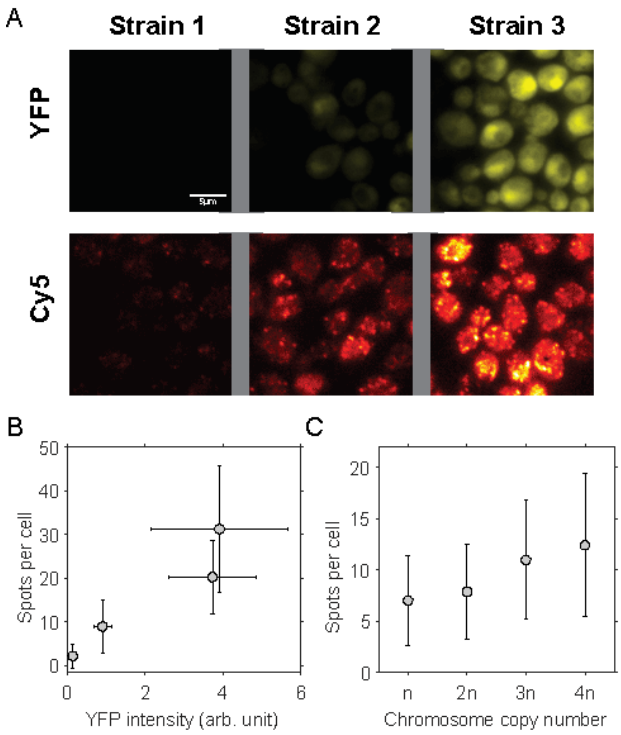


Figure 4.2: Correlation between sFISH spots and protein expression level. **(A)** Fluorescence images of single yeast cells expressing YFP (top row) and sFISH signals from Cy5-labeled probes targeting YFP mRNA (bottom row). Shown from left to right are fluorescence images of the negative control (no YFP expression), low YFP expression, and high YFP expression. Fluorescence intensities in the YFP channel and Cy5 channel are represented by false yellow and red colors, respectively. YFP images are from formaldehyde fixed cells, and Cy5 sFISH images are from methanol fixed cells. **(B)** Correlation plot. The mean number of FISH spots is plotted vs. the mean yEVenus expression level. The error bars are measures of the standard deviation. **(C)** sFISH spots vs. ploidy. sFISH was performed on yeast strains with four different ploidies (1n, 2n, 3n, 4n). The error bars show the standard deviation of the data. The number of spots detected per cell increases monotonically with the number of copies.

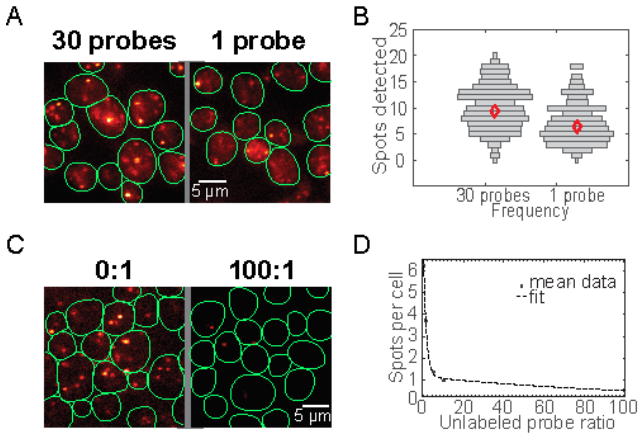


Figure 4.3: Control experiments with KAP104 FISH. **(A)** Comparison between mFISH and sFISH. mFISH and sFISH targeting the KAP104 transcript were performed on the cells grown in the same tube, but fixed with either formaldehyde (mFISH) or methanol (sFISH). Cell boundaries are shown in green. In both cases punctate spots can be seen. mFISH spots are only 2 to 3 times brighter than sFISH spots despite using as many as 48 probes. The relatively low fluorescence signal of mFISH spots compared to sFISH spots is due to the dye (Quasar 670 vs. Cy5) and illumination geometry (epi vs. inclined). **(B)** Spot count distributions from mFISH and sFISH. In these horizontal bar graphs, the position on the vertical axis represents the number of spots per cell, and bar width represents the frequency. Red diamonds are the mean values (9.4 for mFISH and 6.5 for sFISH). The wide distributions represent cell-to-cell variability, not experimental noise. **(C)** sFISH images with (right) and without (left) an unlabeled competitor probe. The ratio above each image is the ratio of unlabeled to labeled probes. **(D)** Mean number of spots per cell vs. unlabeled probe. The spot count decreases monotonically with the concentration of unlabeled probes. The concentration of labeled probes is fixed in the protocol. At 100:1 ratio, only $\sim$0.5 spots per cell are seen, which is similar to the rate of false positives.

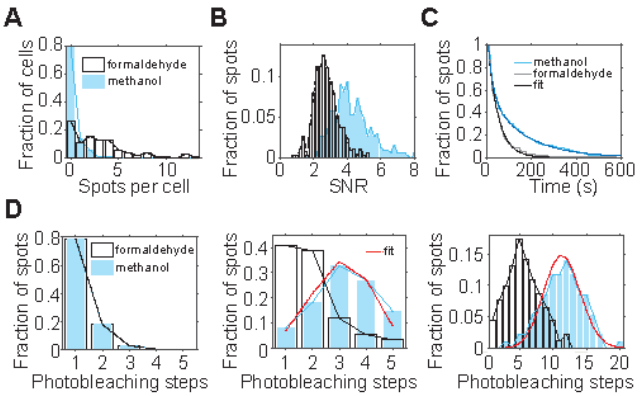


Figure 4.4: Comparison of spot quality between formaldehyde (white) treated samples and methanol (blue) treated samples **(A)** sFISH spots detected from the negative control strain. On average, there is $\sim 0.3$ spots per cell in the methanol treated cells (blue) compared to $\sim 3.1$ spots per cell in the formaldehyde treated sample (white). **(B)** Comparison of signal-to-noise ratio (SNR) of single probes. SNR of a single Cy5 was calculated from fluorescence time traces that captured single-step photobleaching events. Signal is obtained from the single-step drop in fluorescence intensity upon photobleaching, and the noise is calculated as the standard deviation of the signal prior to photobleaching. The histogram shows that the spots from methanol-treated cells (blue) have $\sim$ 2-fold higher SNR than those from formaldehyde-treated cells (white). **(C)** Comparison of Cy5 stability. The population decay curves show that sFISH spots in formaldehyde treated cells photobleach faster than those in methanol-treated cells. **(D)** Comparison of probe number per spot. The number of probes per spot was determined by counting the number of photobleaching steps in the fluorescence time trace. When a single probe was used, most spots photobleached in a single step regardless of the fixative of choice (left). In comparison, when five (middle) or thirty (right) probes targeting the same mRNA were used, more probes were detected from spots in methanol-treated cells than in formaldehyde-treated cells. For the methanol samples treated with multiple probes (middle and right panels), binomial distribution fits are shown in red. For the five-probe experiment (5-probe FISH), Cy5-labeled probes and inclined illumination are used; whereas, for the thirty-probe experiment, Quasar-labeled dyes and epi-illumination are used.

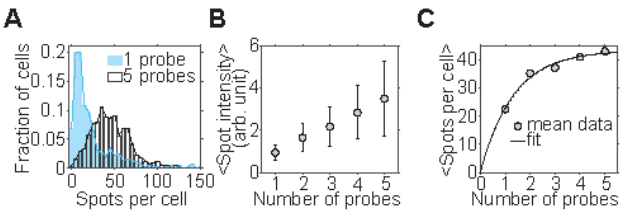


Figure 4.5: Estimation of the hybridization efficiency of single probes. **(A)** The effect of varying the number of probes. The histograms of the number of spots detected per cell are plotted for 1-probe (blue) and 5-probe (white) FISH. **(D)** Spot intensity vs. probe number. The mean spot intensity increases linearly with the number of probes as expected from the binomial distribution. **(C)** Spot number vs. probe number. The mean number of spots detected per cell ($y$) increases with the number of probes ($x$). The fit model is $y = N(1-(1-p)^x)$ where $N$ is the true copy number, and $p$ is the hybridization rate for a single probe. $p$ is extracted to be $53\,\%$.

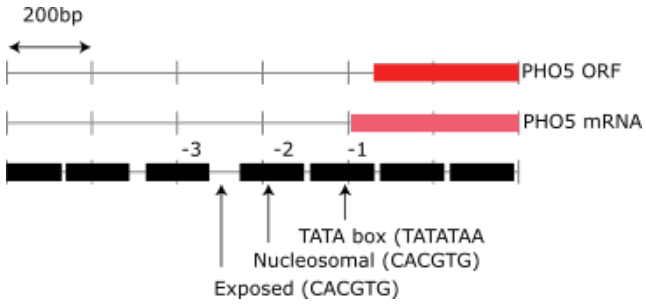


Figure 4.6: PHO5 promoter variants used in this study. PHO5 promoter map. The open reading frame (ORF) and the transcribed region of the PHO5 gene are shown. The reference nucleosome map is retrieved from [71]. The first three nucleosomes are numbered from -1 through -3. The wild-type PHO5 promoter contains two Pho4 binding sites, one in the exposed region between nucleosome -3 and -2, and one within nucleosome -2. The nucleosomal site (CACGTG) has a stronger affinity to Pho4 than the exposed site (CACGTT). The promoter variants used in this study have a common lone high affinity site (CACGTG) in the exposed region with various GC% sequences in the nucleosome -2 region.

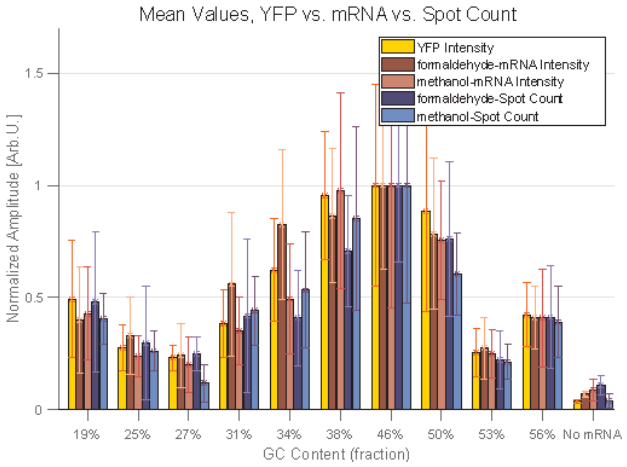


Figure 4.7: YFP Expression Level. YFP intensity is plotted against the percentage of GC in the promoter sequence of the strain. This is a non-monotonic relationship that shows the lowest expression in the $27\%$GC strain and the highest in the $47\%$GC strain. Error bars show the standard deviation of the population which is much larger for strains that are more highly expressing.

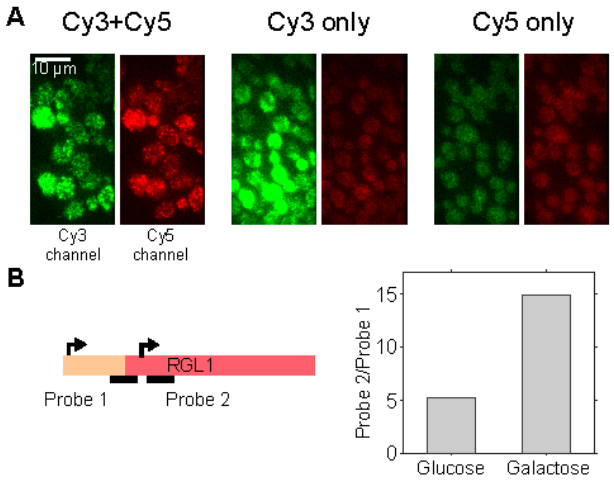


Figure 4.8: Applications of sFISH. **(A)** Demonstration of FRET-FISH in yeast. Fluorescence image acquired under 532-nm excitation was split into two half images based on the emission wavelength. In each image, the green half on the left is from the Cy3 emission channel, and the red half on the right is from the Cy5 emission channel. The images shown represent cells treated with both Cy3- and Cy5-probes (left), Cy3-probe only (middle), and Cy5-probe only (right). Bright, punctate spots were observed in the Cy5 channel only when cells were treated with both probes (left). **(B)** sFISH for mRNA isoform detection. The schematic on the left depicts alternative transcription initiation sites (arrows) at the RGL1 locus, which lead to mRNA isoforms with different lengths. Transcription from the first site produces a full-length mRNA, while from the second site produces a truncated isoform. Using sFISH with two separate probes, the relative fractions of these isoforms can be measured. Probe 1 targets the longer isoform only, whereas probe 2 targets both. The bar plot on the right shows the ratio of sFISH signals with probe 2 to probe 1 measured with glucose (left) or galactose (right) growth media. Here, the mean total fluorescence intensity per cell was used as a proxy for sFISH signal because transcription level was too high to count individual spots.

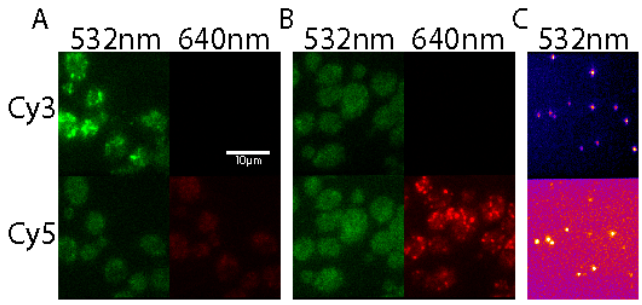


Figure 4.9: Controls for dual-probe FRET-FISH. Each image shown is at the same scale and contrast. A. Cells that have only been treated with a Cy3 labeled probe are shown in dual-view where the emission on the top and bottom panels of a slice were acquired simultaneously. Some bleed-through is expected in FRET experiments. Much of the intensity in the bottom channel during $532\,\mathrm{nM}$ excitation is due to cellular auto-fluorescence, however, there are some peaks in intensity due to Cy3 emission in the Cy5 channel. B. Cells with only Cy5 labeled probes are shown. Under $532\,\mathrm{nM}$ excitation there is very little fluorescence observed. Under $640\,\mathrm{nM}$ excitation, Cy5 is seen in the bottom panel and virtually no emission is observed in the top panel. C. In a TIR setup, Cy3 and Cy5 are observed when labeled on the same DNA oligo bound to the surface with BSA-biotin. These are designed to calibrate the affine transformation that maps the top panel to the bottom panel. They can also be used to calibrate sensitized emission to determine FRET efficiency since these fluorophores are separated by a known number of nucleotides and double stranded DNA is essentially rigid below its persistence length. Variations in spot intensity here are due to non-uniform illumination.

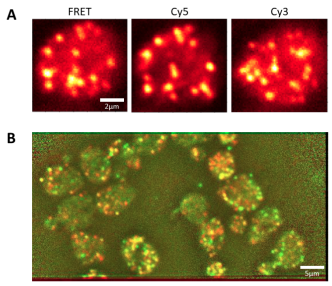


Figure 4.10: Example images of dual-probe FISH-FRET. A. Three channels of a FRET acquisition are shown on a single cell expressing yEvenus mRNA where every target should have both probes. The left and right images are during excitation using $532\,\mathrm{nM}$ light. These images represent Cy5 emission (left) and Cy3 emission (right). Cy5 emission during direct excitation with $640\,\mathrm{nM}$ light is shown in the center. This image was taken second so that the spots that are visible in the FRET (left) image but not the Cy5 (center) image are primarily due to photobleaching. The spots visible in the Cy3 (right) image are due to a combination of non-specific binding (lower melting temperature of the probe) and lower detection efficiency (inactive fluorophores). B. The 256x512 image of Cy5 under direct (Red) and FRET (Green) emission is shown. Where fluorophores were detected in both channels the image is yellow. The direct excitation was measured second. Approximately 80% of fluorophores were detected in both under this excitation condition with the difference due to photobleaching during excitation by FRET.

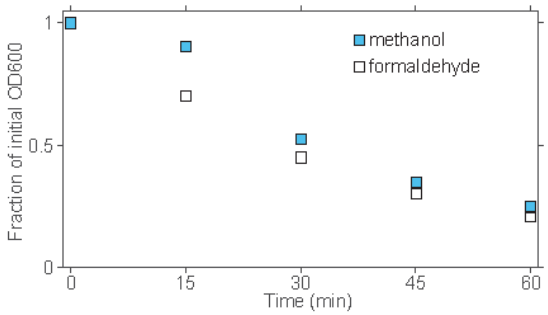


Figure 4.11: Spheroplasting by zymolyase confirmed by absorbance measurement ($OD_{600}$).

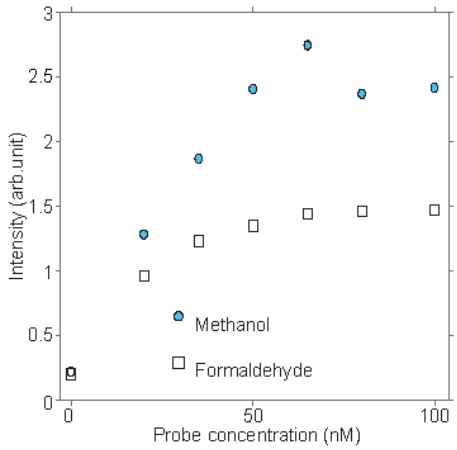


Figure 4.12: sFISH signal vs. probe concentration. sFISH was performed on both methanol- and formaldehyde-fixed cells over a range of probe concentrations. The plot shows that the average fluorescence signal per cell increases with probe concentration and plateaus around 60 nM. Each data point is an average from 200-300 cells. Based on this relationship, the probe concentration of 65 nM is selected for the standard sFISH protocol.

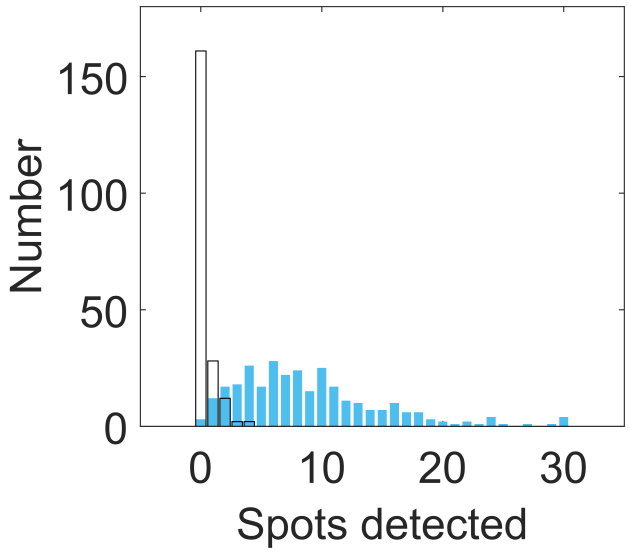


Figure 4.13: False positives in sFISH. Spots counted in each cell are plotted as a histogram. The negative control strain yields a false positive rate of less than 1 per cell (transparent bars). For comparison, the distribution from the low expression strain (positive control) is shown in blue bars.

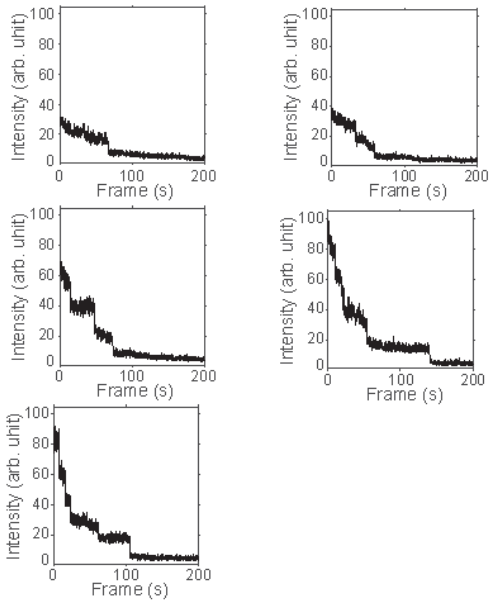


Figure 4.14: Photobleaching of FISH spots. Fluorescence intensities from single spots were monitored under continuous excitation. Most sFISH spots show photobleaching in a single step, (A), which is evidence for the presence of a single fluorophore. Subsequent panels from (B) to (E) show traces taken from 2, 3, 4, and 5 probe treatments, respectively. Overall, the number of photobleaching steps increases with the number of probes used. For these acquisitions, the exposure time was set to 100 ms.

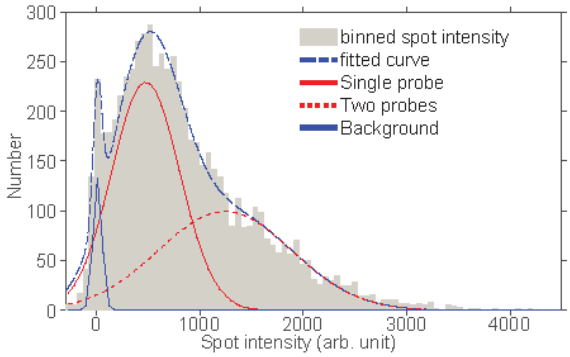


Figure 4.15: Spot ambiguity. Spots inside a cell are not uniformly bright. Using a single fluorophore requires every spot to be considered equally. By integrating the area under the highest intensity Gaussian we can say how often an ambiguous spot occurs. In our lowest expressing cell we find on average one ambiguous spot. We can say that the fit represents background peaks, single fluorophores and ambiguous spots. We find that the rate of ambiguous spots in our highest expressing strain is about 4.4 per cell (13% of spots.)

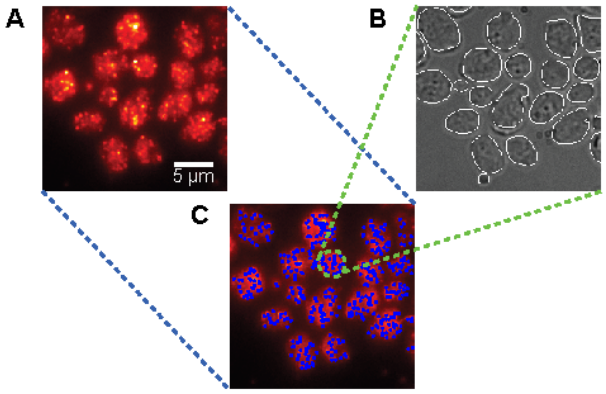


Figure 4.16: Image processing. (A) Raw sFISH data from the Cy5 channel for an mRNA expressing strain. (B) Detection of cell boundaries by applying Sobel filter on the DIC image stack. The local background is approximated by averaging the pixels near the boundary. (C) Spot detection. All local maximum intensity pixels are considered candidate spots. The distribution of the background-subtracted spot intensities exhibits a peak near zero and another peak centered at a higher intensity. Only the spots that belong to the higher intensity peak are qualified as true spots.

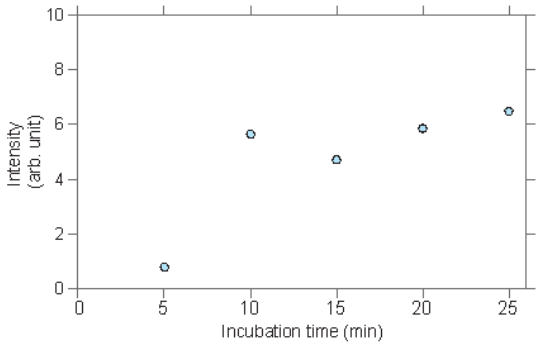


Figure 4.17: sFISH signal vs. zymolyase incubation time. sFISH was performed on cells spheroplasted in zymolyase for different amounts of time. The subsequent probe treatments were identical. As shown in the plot, the average fluorescence intensity per cell plateaus at 10 minutes of incubation. Each data point is an average from 200-300 cells.

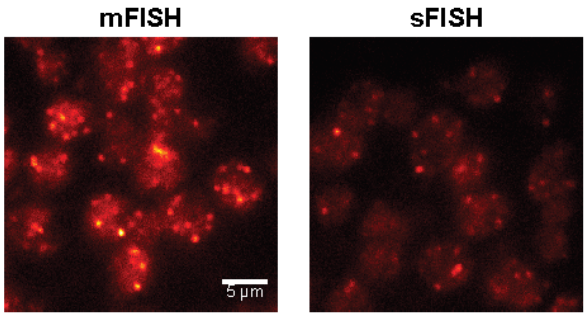


Figure 4.18: Comparison of FISH with five probes and a single probe. Raw FISH images for the low expression yEVenus strain are shown for five Cy5-labeled probes (left) and a single Cy5-labeled probe (right). The average intensity of spots is about three-fold higher when five probes are used.

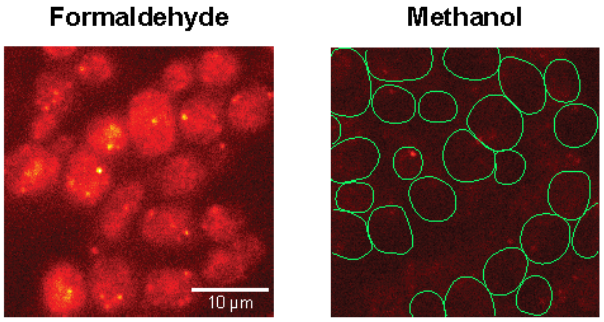


Figure 4.19: Formaldehyde vs. methanol. Negative control cells treated with single probes are shown with formaldehyde fixation (left) and methanol fixation (right). Formaldehyde-fixed cells exhibit higher cellular background as well as more punctate spots (false positives) than methanol-fixed cells. Cell boundaries are shown only for methanol-treated cells.

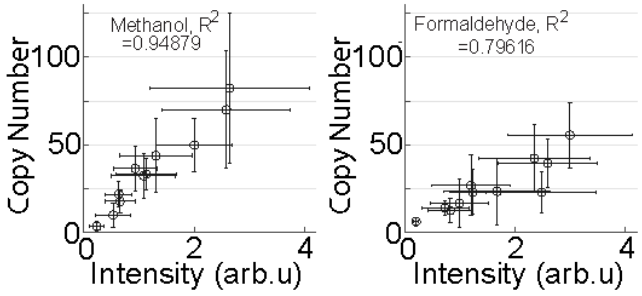


Figure 4.20: Methanol versus Formaldehyde integrated intensity compared to copy number for all strains. The intensity in the Cy5 channel is integrated over the volume of the cell and compared to the number of detected spots. The slope is larger in the case of the methanol treated cells.

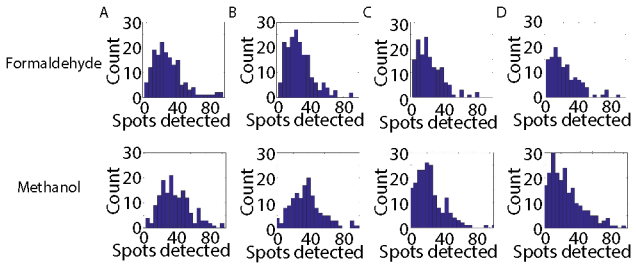


Figure 4.21: Concentration comparison. A. 30x probe set. Formaldehyde and Methanol histograms of detected spots for 65 nM probes. B. 5x 65 nM probes. Histograms are shown for 65 nM working concentration. C. 5x 13 nM probes. Histograms for a working concentration where the total concentration of probes targeting the RNA is 65 nM. D. 65 nM single probe. Histograms of a 65 nM single probe for comparison to multiple probes at the same concentration per probe or the same total concentration per target.

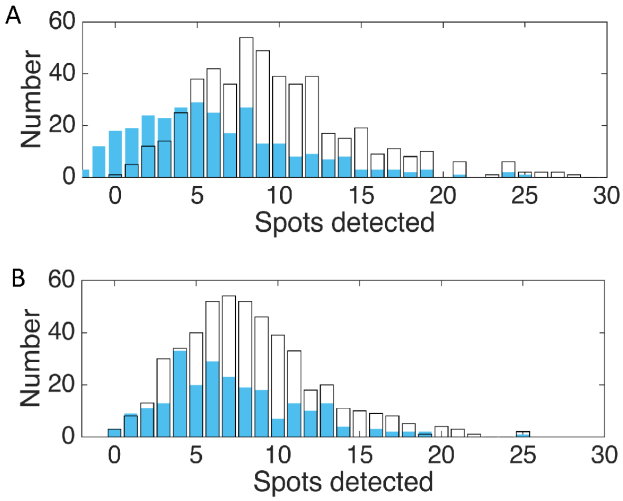


Figure 4.22: Histograms of FISH spot detection. A. yEVenus detection. The multiple probe experiment (white) is compared to the single probe experiment (blue). B. KAP104 detection. The multiple probe experiment (white) is compared to the single probe experiment (blue).

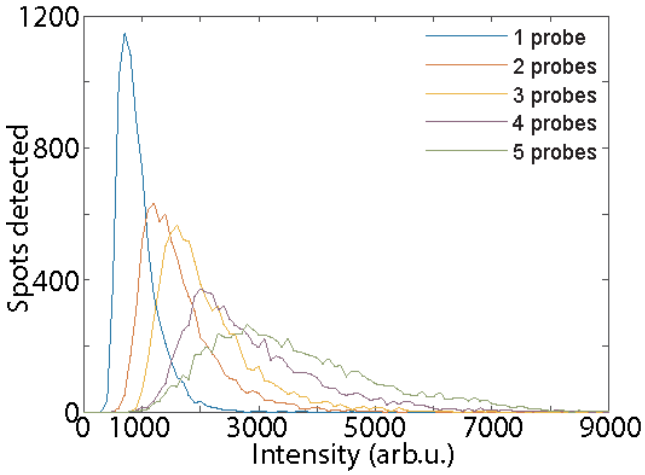


Figure 4.23: Multiple probe intensity. The five probes designed to target yEVenus are used in combination to demonstrate the increase of intensity of the detected spots. There is a linear increase in the mean intensity of the distribution based on the number of probes used.

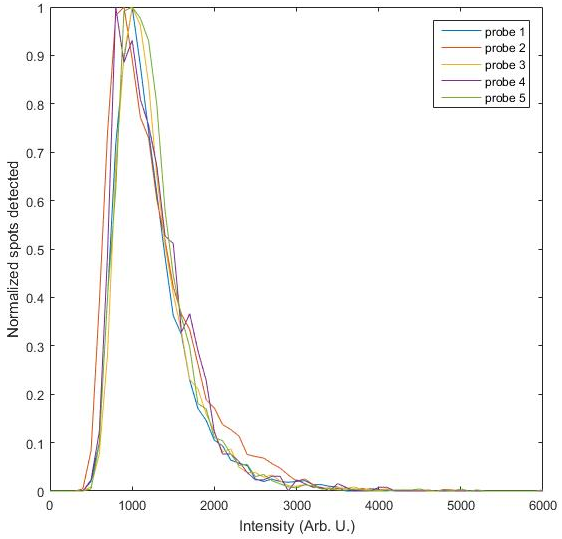


Figure 4.24: Single probe intensities of 5x probe set. Each probe is used in a separate FISH experiment and then the detected intensity of the probes are shown as a histogram. The shift in intensity observed when all five probes are used simultaneously is not present when they are used individually.

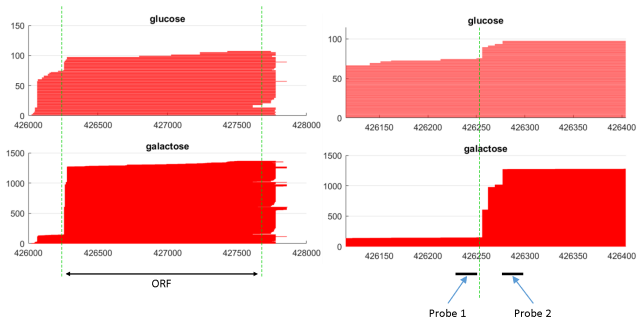


Figure 4.25: Determination of probe location for mRNA isoform profiling. mRNA isoform data for a yeast gene RGL1 (YPL066W) in glucose (top row) and galactose (bottom row) are shown at different zoom levels (zoom-out view on the left column and zoom-in view on the right). The x-axis represents the genomic coordinates around RGL1 on Chromosome XIV. Green vertical lines mark the ORF boundary. mRNA isoforms published in Pelechano et al. [70] are represented by red horizontal lines stacked vertically in the order of start coordinate. As shown, the transcriptional profile of RGL1 changes dramatically from glucose (top row) to galactose (bottom row). The target locations of Probe 1 and Probe 2 are shown as black bars.

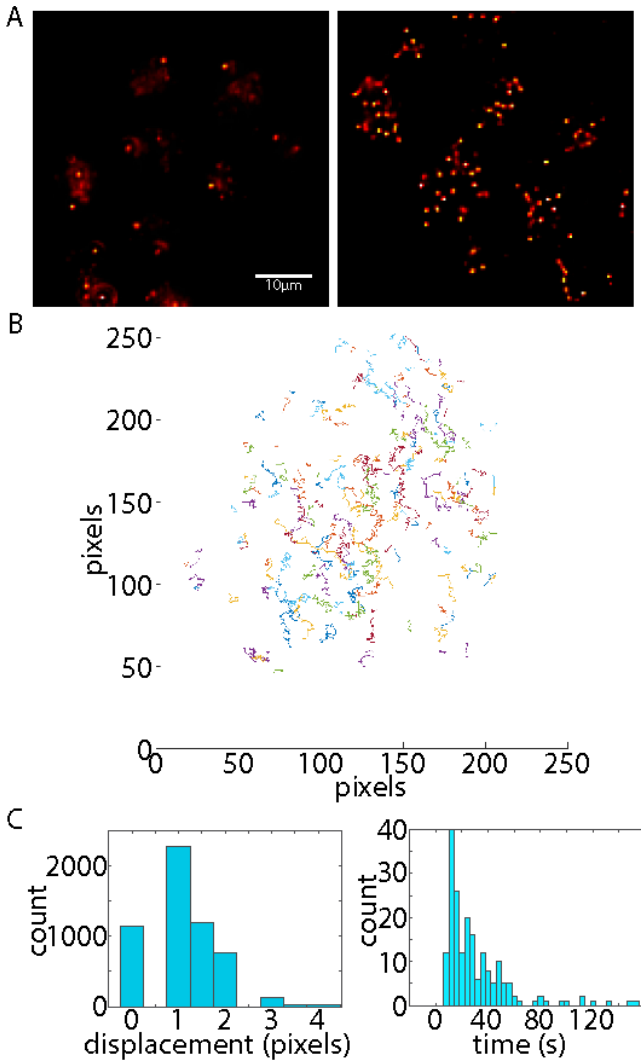


Figure 4.26: DNA paint. A. yEVenus control versus high expression. A time series of 10000 images at $100\,\mathrm{ms}$ is analyzed for peaks. Peak locations in the image are marked and the sum of these observations is used to reconstruct the 2-dimensional localization of fluorophores in the image. It can be seen that there are very few spots, $\sim$1 per cell, in the control that remain in focus for long enough to accumulate a significant signal. B. Trajectories of detected spots. The time trace images are analysed and the trajectory of each detected spot is extracted. This is a representive set of traces. C. Trajectory displacement step size and duration. The displacement of the detected spot in a trace is shown as a histogram (left). The length of each recorded trajectory is shown. Trajectories below 5 steps are excluded.

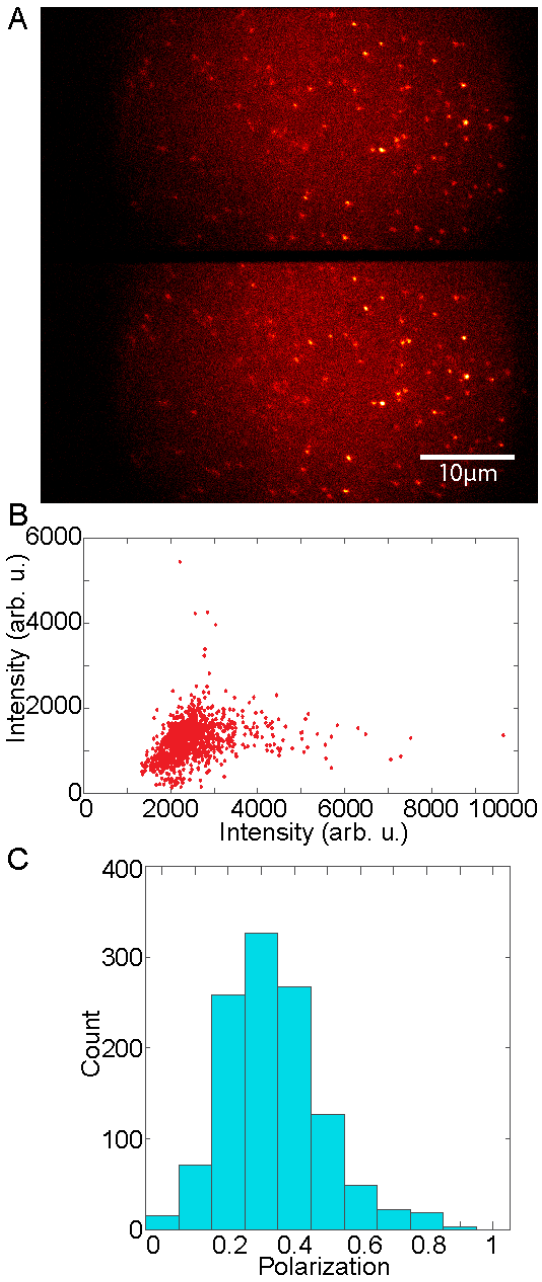


Figure 4.27: Fluorescence anisotropy of single fluorophores. A. Probes bound to surface. The original yEVenus probe contained an internally labeled Cy5 and a biotinylated 3' end. This probe is attached to the surface by flowing bovine serum albumin (BSA) and incubating for 10 minutes with the biotinylated probes. The top channel represents the S polarization and the bottom channel represents the P polarization. B. P versus S intensity. The intensity of single probes in each channel is used to demonstrate that the P polarization is more intense compared to the S polarization for most detected spots. C. Histogram of polarization. The difference of intensities divided by the sum defines the polarization of the spot. The histogram shows that the majority of the spots display some anisotropy with polarization values between (0.2-0.4).

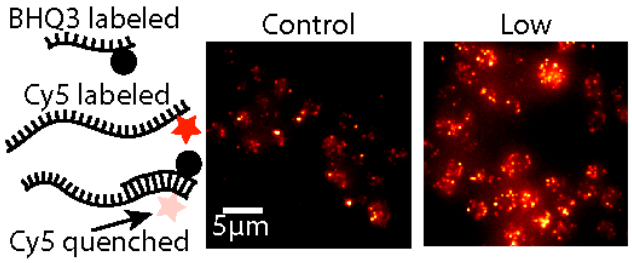


Figure 4.28: Black Hole Quencher probe. BHQ-3 is chosen for the best spectral overlap with Cy5. This quencher functions through Resonant Energy Transfer, but dissipates the energy through thermal energy rather than fluorescence. When the Cy5 probe is bound to the quencher probe, there should be no fluorescent signal. When comparing a negative control to the low expression yEVenus strain, there is still significant signal observed.

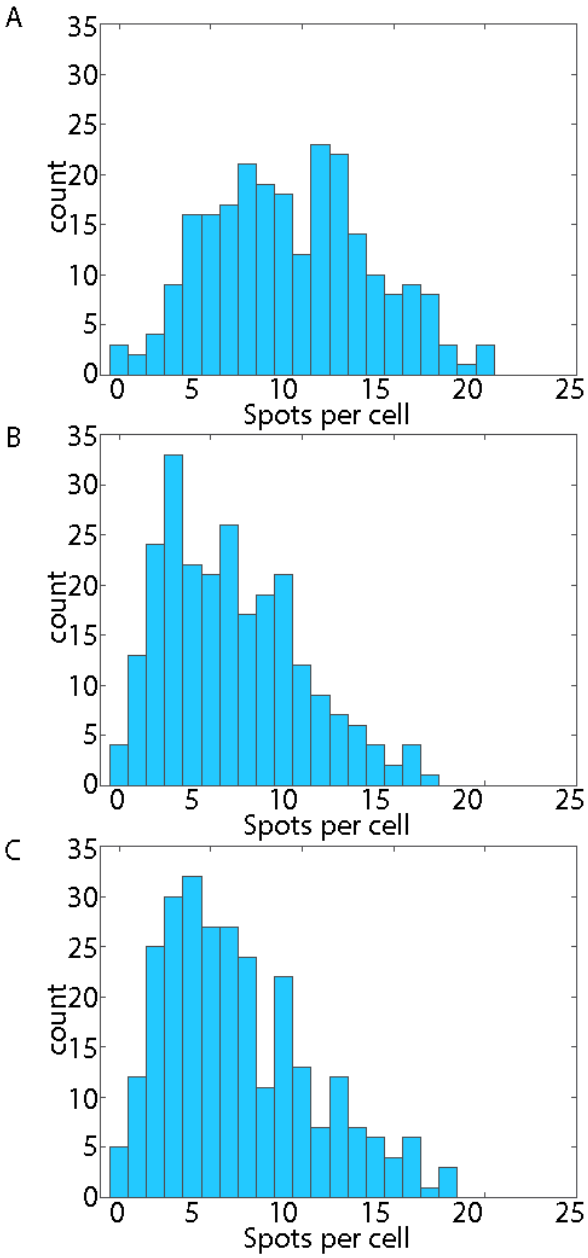


Figure 4.29: End versus internal labeling. A. Multiple end labeled probes. The 30x probe set is used as a reference for the number of spots detected. B. End labeled single probes. The methanol FISH experiment is conducted using an end labeled probe sequence identical to the 26nt sequence used with the internally labeled probe for yEVenus. 5' labeled Cy5 probes were ordered from Eurofins. C. Internally labeled single probes. FISH is performed with an internally labeled Cy5 probe with the fluorophore positioned two nucleotides away from the 5' end.

Table 4.2: DNA sequence for yEVenus probes of similar melting temperature. Probes were designed to have similar GC content and melting temperature. The first probe is used for all of the single probe experiments.

| Melting Temp($^{\circ}C$) | Sequence |
|---|---|
| 47.7 | CA/iCy5/ACCAAAATTGGGACAACACCAGTG/3BioTEG/ |
| 49.7 | TT/iCy5/ACCGTAAGTAGCATCACC |
| 51.1 | TG/iCy5/ACTTTTTCAAGTCTGCCATG |
| 47.8 | AC/iCy5/CGGAGACAGAAAATTTGTGACC |
| 49.2 | CC/iCy5/ATGGAACTGGCAATTTACC |

Table 4.3: Probe sequences. Listed are sequences of probes used against yEVenus mRNA sequence[66]. These sequences are identical to all but two used in a previous study[72].

| Probe Number | Sequence |
|---|---|
| 1 | AATTCTTCACCTTTAGACAT |
| 2 | AATTGGGACAACACCAGTGA |
| 3 | CATCACCATCTAATTCAACC |
| 4 | GACAGAAAATTTGTGACCAT |
| 5 | CATCACCTTCACCTTCACCG |
| 6 | AAGGTCAATTTACCGTAAGT |
| 7 | GTTGGCCATGGAACTGGCAA |
| 8 | ATAACCTAAAGTAGTGACTA |
| 9 | TGTTGTTTCATATGATCTGG |
| 10 | CATGGCAGACTTGAAAAAGT |
| 11 | CTTTCTTGAACATAACCTTC |
| 12 | GTCATCTTTGAAAAAAATAG |
| 13 | CAGCTCTGGTCTTGTAGTTA |
| 14 | GTATCACCTTCAAACTTGAC |
| 15 | TAATTCGATTCTATTAACTA |
| 16 | TCTTCTTTAAAATCAATACC |
| 17 | TTTGTGACCTAAAATGTTAC |
| 18 | GAGAGTTATAGTTGTATTCC |
| 19 | TCAGCAGTGATGTAAACATT |
| 20 | CTTTGATACCATTCTTTTGT |
| 21 | TTGTGTCTAATTTTGAAGTT |
| 22 | TTGAACACCACCATCTTCAA |
| 23 | TTTGTTGATAATGGTCAGCT |
| 24 | GGACCATCACCAATTGGAGT |
| 25 | AATGGTTGTCTGGTAACAAG |
| 26 | AAGGCAGATTGATAGGATAA |
| 27 | CTTTTCGTTTGGATCTTTGG |
| 28 | CTAACAAGACCATGTGGTCT |
| 29 | ATACCAGCAGCAGTAACAAA |
| 30 | ACAATTCATCCATACCATGG |

Table 4.4: Probe sequences. Listed are sequences of probes used against KAP104 mRNA sequence. These sequences are identical to those used in a previous study[68].

| Probe Number | Sequence |
|---|---|
| 1 | GAGTTGCTAGTTGCAACACA |
| 2 | TAGCTTCCATTGCGTTATTA |
| 3 | TCTGAAGATCCTGTAGGGAG |
| 4 | ATTAAATTGTTTCCCCCTAG |
| 5 | TACCTGTCACGTTCGAAACG |
| 6 | CTCTATGCTGCCTATAGTAA |
| 7 | GCATTTGAAGACCAGTTGGA |
| 8 | ATTTGAGGTTAGCTCTAGCA |
| 9 | AAGCCTTAATACTTGGCTCA |
| 10 | AAAATTGAGCGCTGTCTTCC |
| 11 | CTATCCAATAAGGCTTCCAT |
| 12 | CTGAGCGAATCACAGGTGAG |
| 13 | AACTTTGTGTTTGTAGCGGG |
| 14 | TAATGCAAATCTGTGCCCTA |
| 15 | AAACCAGCTTATCTGGTCTG |
| 16 | TTTCCTCATTTACAGTGGTG |
| 17 | TGCAAAAACTCGCAGGCTTC |
| 18 | GTTCAGGAATATTTGGGCTC |
| 19 | GAAGCTTCCAGGAGAACGAT |
| 20 | CTTTATCCTCCAAGAATGCA |
| 21 | TTTTTCACAATACGGGGTGC |
| 22 | TCGTCATTACATCTAGCGTT |
| 23 | ACGCTATATCCATCACTTGA |
| 24 | GCTTCTCTAATAAACCACCT |
| 25 | AATACTTCATTCCACCTTCT |
| 26 | TGGTATTAGTGCTGGTAAGC |
| 27 | ACATGTCATTTTCCTCACTG |
| 28 | GGAGTCGGCGTTTTCAATAA |
| 29 | GGGCGGCAAAATTATTTGCA |
| 30 | TCACTGTCTGACAGCAAAGC |
| 31 | AAGGCTCTGTTGTACACTTC |
| 32 | TTCTCAGGCACTACTATTGT |
| 33 | TCCCTTGAGGGAACAATAGA |
| 34 | TCCTGCAAACATTCTAGCAT |
| 35 | GCAAAGCAGCTTTGTCTTAC |
| 36 | TTCCGTACCAATCAACTTCA |
| 37 | ATCACTATCGTCGTTATGCA |
| 38 | ATCGCATTTATCACAGCAGG |
| 39 | ATCACAGAGCTGTCTACGAT |
| 40 | TTCCGATGGTCACAGACAAG |
| 41 | GGAATCATTGGCAAATGCGC |
| 42 | GCGTTAACGGACAAACACCA |
| 43 | CTACTCTTTTCCTCTACATC |
| 44 | CCGTGCTGGTCAAATTGATA |
| 45 | TCTACATTGCTTGAAAGGCC |
| 46 | GTGTAGATCTCTTGCGCAAA |
| 47 | TTTGGGCAGAATGGTTCATC |
| 48 | AGAAGGAGATTTCATCGGGC |

## OXFORD UNIVERSITY PRESS LICENSE TERMS AND CONDITIONS

Oct 31, 2018

This Agreement between 837 State St ("You") and Oxford University Press ("Oxford University Press") consists of your license details and the terms and conditions provided by Oxford University Press and Copyright Clearance Center.

| | |
|---|---|
| License Number | 4459411335785 |
| License date | Oct 31, 2018 |
| Licensed content publisher | Oxford University Press |
| Licensed content publication | Nucleic Acids Research |
| Licensed content title | mRNA detection in budding yeast with single fluorophores |
| Licensed content author | Wadsworth, Gable M.; Parikh, Rasesh Y. |
| Licensed content date | Jun 28, 2017 |
| Type of Use | Thesis/Dissertation |
| Institution name | |
| Title of your work | Uniquely Quantifying Highly Similar RNA Transcripts at the Single Molecule Level |
| Publisher of your work | Georgia Institute of Technology |
| Expected publication date | Dec 2019 |
| Permissions cost | 0.00 USD |
| Value added tax | 0.00 USD |
| Total | 0.00 USD |
| Title | Uniquely Quantifying Highly Similar RNA Transcripts at the Single Molecule Level |
| Institution name | Georgia Institute of Technology |
| Expected presentation date | Dec 2019 |
| Portions | Portions of Figures 1,2,3,4,5 and text from the results section |
| Requestor Location | 837 State St<br>837 State St<br><br>ATLANTA, GA 30332<br>United States<br>Attn: 837 State St |
| Publisher Tax ID | GB125506730 |
| Billing Type | Invoice |
| Billing Address | 837 State St<br>837 State St<br><br>ATLANTA, GA 30332<br>United States<br>Attn: 837 State St |
| Total | 0.00 USD |

bio-protocol

**Bio-Protocol**
**License to Publish**

The following License must be signed and returned to Bio-protocol before a manuscript can be accepted for publication. If the copyright in the contribution is owned by an entity other than the author (such as the author's employer), the copyright owner or its authorized representative must co-sign this form.

This License to Publish (**"License"**) is entered into as of 04/11, 2018 (the **"Effective Date"**) by and between Bio-protocol LLC (**"Bio-protocol"**) and the person(s) identified as an author below (individually and collectively, **"Author"**). This License applies to the Author's protocol or article currently titled Single-probe RNA FISH in Yeast

and all associated supplemental materials, data, images, audio and/or video (collectively, the **"Article"**), which the Author is submitting for publication in one of Bio-protocol's websites.

1. **License Grant.** Subject to the terms of this License, Author hereby grants to Bio-protocol the sole, exclusive, irrevocable, worldwide and royalty-free right to publish, reproduce, distribute, transmit, publicly display, store, translate, create derivative works from and otherwise use or exploit the Article in any form, manner, format or medium, whether known or hereafter developed, in any language, for the entire duration of any such right and any renewal or extension thereof and to permit and sublicense others through multiple tiers to do any or all of the foregoing.

2. **Reservation of Rights.**
   a. Ownership. Author reserves ownership of the Article and all of Author's other proprietary rights with respect to the Article, including the copyright and any patent rights Author may have in any inventions described in the Article, subject to the rights granted to Bio-protocol in Section 1. Bio-protocol shall be the holder of title for the purpose of copyright registration.
   b. Other Rights. Author retains the following non-exclusive rights to use the Article without further permission after the publication of the Article by Bio-protocol, provided that Author will reference (and, where applicable, require any third party to reference) the appropriate citation of the publication by Bio-protocol of the Article when exercising any of the rights retained by the Author in this subsection 2.b.:
      i. Reprint the Article in print collections of Author's own writings;
      ii. Reprint the Article in print format for inclusion in a thesis or dissertation that the Author writes for a degree granted by an educational institution;
      iii. Present the Article orally;
      iv. Reproduce the Article for use in courses that the Author is teaching (if the Author is employed by an academic institution, that institution may also reproduce the Article for course teaching at that institution);
      v. Distribute photocopies of the Article to colleagues for non-commercial purposes only;
      vi. Reuse figures or tables in the Article for future publications written by the Author; and

   Author agrees to obtain prior consent from Bio-protocol for any uses not expressly authorized in this Section 2.

3. **Enforcement.** Author will promptly notify Bio-protocol if Author becomes aware of any known or suspected infringement of the copyright in the Article. Bio-protocol will be entitled to enforce the copyright against any third-party infringer. If Bio-protocol asserts or files any claim, suit, or action (a **"Claim"**) against any such third-party infringer, Author will cooperate with Bio-protocol, at Bio-protocol's request, in the prosecution and settlement of such Claim, including joining as a party to such suit or action where reasonably required.

4. **Warranties.** Author represents and warrants that:

a. Author has all rights and licenses necessary to grant to Bio-protocol the rights in the Article specified in Section 1 above;

b. The Article has not been published anywhere else and Bio-protocol's use of the Article in compliance with the license granted in Section 1 will not infringe, misappropriate or violate any third party's copyright, proprietary or personal rights, or any applicable law or regulation;
c. the Article is original, and that Author has made all reasonable efforts to ensure the accuracy of any factual information contained in the Article; and
d. Author's agreement to the terms and conditions of this License will not violate or conflict with any obligations owed by Author to a third party.

5. **Indemnification.** Author will indemnify and hold harmless Bio-protocol and its affiliates, employees, directors, officers, agents, customers and sublicensees against any and all claims, liabilities, damages, costs, and expenses directly or indirectly arising from or relating to any breach of or inaccuracy in any of representations or warranties made by Author under this License.

6. **General.** This License is governed by and construed in accordance with the laws of the State of California, excluding that body of law known as conflict of laws. Any legal action or proceeding arising under this License will be brought exclusively in the federal or state courts located in the Northern District of California and the parties hereby irrevocably consent to the personal jurisdiction and venue therein. Bio-protocol may freely assign or transfer this License, in whole or in part, by operation of law or otherwise. This License will bind and inure to the benefit of each party's successors and permitted assigns. If, for any reason, a court of competent jurisdiction finds any provision of this License invalid or unenforceable, that provision of the License will be enforced to the maximum extent permissible and the other provisions of this License will remain in full force and effect. The failure by either party to enforce any provision of this License will not constitute a waiver of future enforcement of that or any other provision. This License constitutes the complete and exclusive understanding and agreement between the parties regarding its subject matter and supersedes all prior or contemporaneous agreements or understandings, written or oral, relating to its subject matter. Any waiver, modification or amendment of any provision of this License will be effective only if in writing and signed by duly authorized representatives of both parties. This License may be executed in counterparts, each of which will be deemed an original, but all of which together will constitute one and the same instrument. In the event that Bio-protocol does not publish the Article in any medium, within two (2) years of the Effective Date, this License shall be null and void. Bio-protocol recognizes that an Article prepared under certain governmental grants or contracts may be subject to the government's prior non-exclusive rights to use the Article for non-commercial, governmental purposes to the extent provided in the contract or grant or other regulations.

If the Article was prepared jointly by multiple authors, please identify all co-authors below. If any co-authors or other copyright holders of the Article are not personally signing this License (each, a "**Non-Signing Author**"), the person designated below as the Corresponding Author represents and warrants that he or she (1) has informed all the Non-Signing Authors of the terms of this License, (2) has the authority to act as the agent for the Non-Signing Authors in relation to this License, and (3) is signing this License on behalf of all the Non-Signing Authors and also on his or her own behalf. The signatory must inform Bio-protocol of any changes in authorship.

bio-protocol

Corresponding Author
Name: Harold Kim Signature: _____ Date: 5/1/18

☒ Corresponding Author signing on behalf of all Non-Signing Authors.

Author
Name: Gable Wadsworth Signature: _____ Date: 5/1/18
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

bio-protocol

**Bio-Protocol**
**License to Publish**

The following License must be signed and returned to Bio-protocol before a manuscript can be accepted for publication. If the copyright in the contribution is owned by an entity other than the author (such as the author's employer), the copyright owner or its authorized representative must co-sign this form.

This License to Publish ("**License**") is entered into as of 05/01, 2018 (the "**Effective Date**") by and between Bio-protocol LLC ("**Bio-protocol**") and the person(s) identified as an author below (individually and collectively, "**Author**"). This License applies to the Author's protocol or article currently titled Dual-probe RNA FRET-FISH in Yeast

and all associated supplemental materials, data, images, audio and/or video (collectively, the "**Article**"), which the Author is submitting for publication in one of Bio-protocol's websites.

1. **License Grant.** Subject to the terms of this License, Author hereby grants to Bio-protocol the sole, exclusive, irrevocable, worldwide and royalty-free right to publish, reproduce, distribute, transmit, publicly display, store, translate, create derivative works from and otherwise use or exploit the Article in any form, manner, format or medium, whether known or hereafter developed, in any language, for the entire duration of any such right and any renewal or extension thereof and to permit and sublicense others through multiple tiers to do any or all of the foregoing.

2. **Reservation of Rights.**
   a. Ownership. Author reserves ownership of the Article and all of Author's other proprietary rights with respect to the Article, including the copyright and any patent rights Author may have in any inventions described in the Article, subject to the rights granted to Bio-protocol in Section 1. Bio-protocol shall be the holder of title for the purpose of copyright registration.
   b. Other Rights. Author retains the following non-exclusive rights to use the Article without further permission after the publication of the Article by Bio-protocol, provided that Author will reference (and, where applicable, require any third party to reference) the appropriate citation of the publication by Bio-protocol of the Article when exercising any of the rights retained by the Author in this subsection 2.b.:
      i. Reprint the Article in print collections of Author's own writings;
      ii. Reprint the Article in print format for inclusion in a thesis or dissertation that the Author writes for a degree granted by an educational institution;
      iii. Present the Article orally;
      iv. Reproduce the Article for use in courses that the Author is teaching (if the Author is employed by an academic institution, that institution may also reproduce the Article for course teaching at that institution);
      v. Distribute photocopies of the Article to colleagues for non-commercial purposes only;
      vi. Reuse figures or tables in the Article for future publications written by the Author; and

   Author agrees to obtain prior consent from Bio-protocol for any uses not expressly authorized in this Section 2.

3. **Enforcement.** Author will promptly notify Bio-protocol if Author becomes aware of any known or suspected infringement of the copyright in the Article. Bio-protocol will be entitled to enforce the copyright against any third-party infringer. If Bio-protocol asserts or files any claim, suit, or action (a "**Claim**") against any such third-party infringer, Author will cooperate with Bio-protocol, at Bio-protocol's request, in the prosecution and settlement of such Claim, including joining as a party to such suit or action where reasonably required.

4. **Warranties.** Author represents and warrants that:

a. Author has all rights and licenses necessary to grant to Bio-protocol the rights in the Article specified in Section 1 above;

b. The Article has not been published anywhere else and Bio-protocol's use of the Article in compliance with the license granted in Section 1 will not infringe, misappropriate or violate any third party's copyright, proprietary or personal rights, or any applicable law or regulation;
c. the Article is original, and that Author has made all reasonable efforts to ensure the accuracy of any factual information contained in the Article; and
d. Author's agreement to the terms and conditions of this License will not violate or conflict with any obligations owed by Author to a third party.

5. **Indemnification.** Author will indemnify and hold harmless Bio-protocol and its affiliates, employees, directors, officers, agents, customers and sublicensees against any and all claims, liabilities, damages, costs, and expenses directly or indirectly arising from or relating to any breach of or inaccuracy in any of representations or warranties made by Author under this License.

6. **General.** This License is governed by and construed in accordance with the laws of the State of California, excluding that body of law known as conflict of laws. Any legal action or proceeding arising under this License will be brought exclusively in the federal or state courts located in the Northern District of California and the parties hereby irrevocably consent to the personal jurisdiction and venue therein. Bio-protocol may freely assign or transfer this License, in whole or in part, by operation of law or otherwise. This License will bind and inure to the benefit of each party's successors and permitted assigns. If, for any reason, a court of competent jurisdiction finds any provision of this License invalid or unenforceable, that provision of the License will be enforced to the maximum extent permissible and the other provisions of this License will remain in full force and effect. The failure by either party to enforce any provision of this License will not constitute a waiver of future enforcement of that or any other provision. This License constitutes the complete and exclusive understanding and agreement between the parties regarding its subject matter and supersedes all prior or contemporaneous agreements or understandings, written or oral, relating to its subject matter. Any waiver, modification or amendment of any provision of this License will be effective only if in writing and signed by duly authorized representatives of both parties. This License may be executed in counterparts, each of which will be deemed an original, but all of which together will constitute one and the same instrument. In the event that Bio-protocol does not publish the Article in any medium, within two (2) years of the Effective Date, this License shall be null and void. Bio-protocol recognizes that an Article prepared under certain governmental grants or contracts may be subject to the government's prior non-exclusive rights to use the Article for non-commercial, governmental purposes to the extent provided in the contract or grant or other regulations.

If the Article was prepared jointly by multiple authors, please identify all co-authors below. If any co-authors or other copyright holders of the Article are not personally signing this License (each, a "**Non-Signing Author**"), the person designated below as the Corresponding Author represents and warrants that he or she (1) has informed all the Non-Signing Authors of the terms of this License, (2) has the authority to act as the agent for the Non-Signing Authors in relation to this License, and (3) is signing this License on behalf of all the Non-Signing Authors and also on his or her own behalf. The signatory must inform Bio-protocol of any changes in authorship.

bio-protocol

Corresponding Author
Name: Harold Kim Signature: _____ Date: 5/1/18
☒ Corresponding Author signing on behalf of all Non-Signing Authors.

Author
Name: Gable Wadsworth Signature: _____ Date: 5/1/18
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

Author
Name: _____ Signature: _____ Date: _____
(optional if the box for Corresponding Author is checked)

Table 4.5: DNA sequence for KAP104 probes used in sFISH experiments.

| Melting Temp($^{\circ}C$) | Sequence |
|---|---|
| 56.0 | CA/iCy5/CAGGTGAGAAATTTGGATTCGA |
| 55.4 | CACAGGTGAGAAATTTGGATTCGA |

Table 4.6: DNA sequence for RGL1 probes.

| | Melting Temp($^{\circ}C$) | Sequence |
|---|---|---|
| Probe 1 | 60.2 | [Cy5]TATAACTGGGTGCGTCATTTTCACTTCTTCG |
| Probe 2 | 68.3 | [Cy5]TAACGTATCAGGTAAACCGGGGCAGCCCCGT |

# CHAPTER 5

# LOCALIZATION AND DIFFUSION OF TLC1 TRANSCRIPTS

The following work represents unpublished work in collaboration with Virginia Zakian and her student Patricia Daniela Garcia at Princeton. The work that will appear in a future publication is presented here with their consent.

The regulation of non-coding RNA transcripts is of fundamental importance in the cell cycle. The life cycle of these transcripts can be very complex and different compared to the more studied messenger RNA transcripts. Since many of these transcripts fulfill functions critical to cell viability, perturbations to the systems surrounding these transcripts can give rise to variations in expression and localization patterns of these transcripts. One change to a protein can completely alter the ability of the non coding RNA to perform its function. One such non-coding RNA is TLC1 which is the RNA sub unit of telomerase and binds POP proteins which are also components of RNase P and RNase MRP[73]. During its life cycle the TLC1 transcript must undergo transport to the nucleolus for pop protein binding, transport into the cytoplasm for the remaining proteins to associate and complete the process to convert it into the mature telomerase, and then import back to the nucleus to perform its function. The multi-step transport process involved in producing telomerase for the cell means that there are many opportunities for the process to break down. In order to understand any change in this process it is important to use a method that can give spatial information. One popular method to measure RNA transcripts is single molecule Fluorescent In Situ Hybridization(smFISH)[63]. This method has been used to determine models of stochastic gene expression and quantify transcripts at the single cell and single molecule level in a wide variety of ways. However, since the majority of transcripts studied with smFISH have been mRNA there is not much reason to consider transport across the nuclear membrane. Transcripts are mostly either nascent or in the cytoplasm to produce

proteins. Another challenge is that targeting a transcript that is coordinated with proteins is likely more difficult with a multiple probe labeling strategy. Here we demonstrate the utility of our previously reported single fluorophore FISH method[59] to detect and distinguish organelle level distributions of TLC1 transcripts. Furthermore, we examine the statistics of the spatial distributions of these transcripts to determine how TLC1 mislocalizes under mutation of Pop1 and Pop6 proteins.

## 5.1 Mutation dependent localization of TLC1

### 5.1.1 Results

TLC1 localization is determined by RNA FISH following a single-probe protocol[59]. The 52-nt long probe targeting TLC1 includes a short 10-nt toehold, a 16-nt stem region and a 10-nt loop of random sequence (Figure 5.1). The hairpin design was adopted to increase probe specificity. The nucleolus was marked with another FISH probe targeting ITS1 RNA, and the nucleus was stained with DAPI. We analyzed three-dimensional localization of TLC1 spots with reference to Cy5 and DAPI signals. TLC1 spots that colocalize with Cy5 and DAPI are categorized as nucleolar and nuclear, respectively. Spots that colocalize with neither are categorized as cytoplasmic. Examples images of the wild-type, pop1 and pop6 strains for both permissive growth ($24\,^{\circ}\mathrm{C}$) and semi-permissive growth ($30\,^{\circ}\mathrm{C}$) are shown in Fig. 5.2. Both TLC1 and ITS1 FISH signals vanished in the presence of 10-fold excess unlabeled competitor probes Fig. 5.3, indicative of the probe specificity. We detected $8.6 \pm 0.3$ FISH spots per cell for the wild-type strain, $24.9 \pm 0.8$ for the pop1 mutant, and $16.5 \pm 0.7$ for the pop6 mutant at $24\,^{\circ}\mathrm{C}$. This 2- to 3-fold difference in spot number between the wild-type and pop mutant strains is consistent with RT-qPCR measurements at $24\,^{\circ}\mathrm{C}$. For each strain, the spot number was higher at $30\,^{\circ}\mathrm{C}$ than at $24\,^{\circ}\mathrm{C}$ growth (Figure 5.1B). At the single-cell level, the spot number was highly variable. For the wild-type grown at $24\,^{\circ}\mathrm{C}$, this variability was mostly due to the spots in the nucleus, but for all other cases, the spots in the cytoplasm dominated the variability (Fig. 5.6). We quantified the subcellular

distribution of TLC1 spots in terms of the cytoplasmic fraction ($f_c$), the ratio of cytoplasmic to total spot count, and the nucleolar subfraction ($f_{no}$), the ratio of nucleolar to nuclear spot count. If TLC1 transcripts were uniformly distributed throughout the cell, $f_c$ would be simply equal to the cytoplasmic volume fraction of each cell. However, the histograms of $f_c$ obtained at the single cell level (white columns, Figure 5.1D) are markedly more disperse than the histogram of cytoplasmic volume fractions (Figure 5.1C), which reveals that TLC1 localization is highly regulated. At $24\,^{\circ}\mathrm{C}$ growth, the histograms also exhibit strong bimodality with well-separated peaks near $f_c = 0$ (nuclear) and $f_c = 1$ (cytoplasmic). Our morphology-based cell-cycle analysis shows that the first peak is high in the G1 phase while the second peak is high in the S and G2/M phases (Fig. 5.5). In comparison to the wild-type strain, pop mutants show a lower frequency at $f_c = 0$ and higher frequency at intermediate values of $f_c$ (Figure 5.1D), which indicates that pop mutation leads to more frequent localization of TLC1 transcripts to the cytoplasm. At $30\,^{\circ}\mathrm{C}$ growth, the distribution is no longer bimodal (blue columns, Figure 5.1D). Instead, pop6 mutant exhibits a distribution similar to that of the cytoplasmic volume fraction (Figure 5.1C), which suggests that pop6 mutation compromises active nuclear localization of TLC1 at $30\,^{\circ}\mathrm{C}$ growth. A similar temperature and mutation dependence is observed for nucleolar localization of TLC1 inside the nucleus. At $24\,^{\circ}\mathrm{C}$ growth, the wild-type strains show nuclear spots mostly outside the nucleolus, as reflected by the large population near $f_{no} = 0$ (white columns, Figure 5.1D). Pop mutation leads to a moderate decrease in nucleolar exclusion. Nucleolar localization becomes substantially more frequent at $30\,^{\circ}\mathrm{C}$ growth (blue columns, Figure 5.1E). However, under no circumstances, complete nucleolar localization ($f_c = 0$ and $f_{no} = 1$) was observed (Fig. 5.4B).

### 5.1.2 Methods

FISH probes were designed based on a short 26-nt region of either TLC1 or ITS1 RNA and contain a 10-nt overhang that induces hairpin formation according to mFOLD[74]

Table 5.1: TLC1 and ITS1 probe sequences. Probe sequences for the hairpin probes for TLC1 and ITS1 and their competitors are shown.

| Melting Temp($^{\circ}C$) | Sequence | |
|---|---|---|
| TLC1 | 47.7 | /5Cy3/GGATCGGTACGAAGAAGGAATAACAGTAGAAAGGTACTGTTATTCCTTCTTC |
| TLC1 control | 49.7 | GGATCGGTACGAAGAAGGAATAACAG |
| ITS1 | 51.1 | /5Cy5/GCCCGGCTGGACTCTCCATCTCTTGTCTTCAGATTGAACGGAAGACAAGAGATGGAGAGT |
| ITS1 control | 47.8 | GCCCGGCTGGACTCTCCATCTCTTGTCTTC |

(Table 5.1). Probes were reconstituted in TE buffer and an annealing cycle is run on a PCR machine before use.

Cells were grown overnight in $150\,\mathrm{mL}$ of SD Complete or SD -ura dropout media at either $24\,^{\circ}\mathrm{C}$ or $30\,^{\circ}\mathrm{C}$. Besides the following modifications, cells were prepared for FISH as described previously[59]. Cells were harvested once 0.6 $OD_{600}$ was reached. Cells were fixed in 10% v/v formaldehyde at room temperature for $45\,\mathrm{min}$ and spheroplasted and permeabilized overnight in ethanol. Hybridization was performed overnight with a working concentration of $50\,\mathrm{nM}$ for both probes. Cells were pelleted and resuspended in $1\,\mathrm{mL}$ of wash buffer containing $1\,\mu\mathrm{L}$ of $1\,\mathrm{mg/mL}$ DAPI. Cells were incubated in this wash buffer for $10\,\mathrm{min}$ at $30\,^{\circ}\mathrm{C}$ to wash out weakly-bound probes. Cells were washed once more without incubation and then aspirated and kept at $4\,^{\circ}\mathrm{C}$ until imaged. $3\,\mu\mathrm{L}$ of concentrated cells were applied to a slide pre-cleaned with ethanol and were mixed with $3\,\mu\mathrm{L}$ of oxygen scavenging buffer by gentle pipetting. Slides were sealed with epoxy. Z-stack images were acquired at $100\,\mathrm{ms}$ exposure by Micromanager[75], in $200\,\mathrm{nm}$ steps over $10\,\mu\mathrm{m}$ using a motorized stage. Laser output was set to produce $25\,\mathrm{mW}$ of power at the sample plane for both $640\,\mathrm{nm}$ and $532\,\mathrm{nm}$ illumination (1185055, Coherent; LCX-532L-100, Oxxius), and the LED controller was set to $0.5\,\mathrm{V}$for UV illumination (M375L4, Thorlabs). The laser light was spun by a 2D-galvo system to achieve a uniform illumination[38]. Cell segmentation and FISH spot identification were performed using Matlab image processing toolbox. Nucleolus and nucleus locations were determined by intensity-based thresholding. Objects of volume smaller than 10 pixels were removed from the analysis. Spot location was extracted by applying a CLEAN algorithm[76]. In this algorithm, a Gaussian profile

fit to a maximum intensity spot is sequentially subtracted from the original image until the intensity in the cell falls below the minimum acceptable intensity of a single fluorophore. Cell cycle assignment to G2/M was made based on the eccentricity of the nucleus. Cell cycle assignment to G2/M was made by separating the bimodal distribution of the nucleus size: Cells with smaller nuclei were assigned to G1 phase, and cells with larger nuclei were assigned to S phase.

## 5.2 Mutation dependent diffusion of TLC1

In order to become the mature holoenzyme telomerase, a TLC1 transcript must undergo a series of transport, folding, and binding steps (Figure 5.8C). This involves transport into the nucleolus, which does not have a membrane and makes up about 25% of the volume inside the nuclear membrane. The nucleolus is composed of granular and fibrilar regions around a cavity and condenses around the DNA that codes for ribosomal components[77]. The fibrilar regions contain the rDNA and both regions are composed of ribonucleoproteins that are involved in the production of ribosomes. The TLC1 transcript enters the nucleolus because it is the location of POP1 and POP6 proteins which are subcomponents of RNase P and RNase MRP. As such these proteins are most concentrated in the nucleolus. Once the Pop proteins are bound, the TLC1 transcript exits the nucleolus and then is transported through the nuclear membrane to bind Ku and Est proteins, which are located in the cytoplasm. After binding Ku and Est proteins, the TLC1 transcript is the mature holoenzyme and is transported back into the nucleus to maintain telomeres. The telomeres in yeast are attached to the nuclear membrane so that the majority of TLC1 transcripts should be found near the nuclear membrane[78] (Figure 5.10A). The location of this membrane is then the most important thing to determine with confidence. It is unlikely that the location of the membrane could be determined by a depletion in intensity due to it being smaller than the diffraction limit at about $20\,\mathrm{nm}$ thick[79].

The diffusion of TLC1 transcripts is determined by projecting all detected spots onto the

axis of minimum distance to the nuclear membrane. This is determined by creating a vector of all the boundary pixels and calculating the minimum euclidean distance between this set of points and the sub pixel location of a spot determined by a 3-dimensional Gaussian fit. The volume of the nucleus and nucleolus are calculated based on an intensity threshold used to logically select the portion of the image to consider the nucleus or nucleolus. In the case of the nucleus, the intensity cutoff was chosen by observing that there seemed to be two peaks in the volume of nuclei in the images when an arbitrary cutoff was chosen. This intensity cutoff was modified to stretch the size of the first peak away from zero until it appeared Gaussian. When comparing this selection of nucleus size to the relative localization of TLC1 transcripts they both appeared to be at approximately a radius of 7-10 pixels away from the centroid of the intensity distribution of the nucleus (Figure 5.9B-C). Furthermore, the size of the nucleus relative to the total size of the cell is similar to literature measurements of the nucleus size ($\sim$15%)[80] and the same goes for the nucleolus size ($\sim 25\%$)[81].

It is observed that the diffusion of TLC1 under can be reasonably well described by Gaussian fit of the distance away from the membrane, which would be expected for Brownian motion (Figure 5.11). There are some transcripts in that seem better characterized by an additional Gaussian term. The additional term suggests that there is an change in the diffusion of some of these terms, which could imply an active process of diffusion. In the mutants and at elevated temperature the diffusion of transcripts changes and in all cases the transcripts are super diffusive compared to the TLC1 transcripts in the wild-type strain. This diffusion is likely because of the need for the transcript to recruit the Ku and Est proteins in the cytoplasm so they are most likely actively transported back across the nuclear membrane only once they acquire these proteins and are able to fold into the mature form of telomerase.

## 5.3 Two state model of TLC1 transcription

The two state model describes transcription when the gene can be turned on and off. This model is able to account for distributions of transcripts that do not match a Poisson distribution, which would represent the simple case where there is no secondary regulation of transcription. In particular the two state model was necessary to describe a case where the number of transcripts is dependent on bursts of active transcription with long periods of gene inactivation. The shape of the distribution is then characterized by burst size and burst frequency, which are represented by the rate of transcription $s$ divided by the rate of gene inactivation $k_{off}$ and the rate of gene activation $k_{on}$ divided by the rate of mRNA decay $\gamma$. These parameters can be either fit by a gamma function or the full analytical solution for which the gamma function is the case where the the rate of gene inactivation is significantly larger than the rate of gene activation and somewhat larger than the rate of mRNA decay. Gene expression in this case is characterized by short infrequent bursts of transcription.

To establish that the two state model is applicable it is necessary to confirm that there are nascent sites of transcription detected in the population of cells. This is not an easy task. There is no reason that it can immediately be assumed that an intense spot in the nucleus is actually a nascent transcription site rather than the overlap of the fluorescent signal from multiple telomerase enzymes being localized to a telomere. In order to determine whether any of the spots detected could be transcription sites, the intensity of each spot is ploted on the axis of the minimum distance to the nuclear membrane (Figure 5.12). It can be seen that the majority of intense spots occur two pixels or further away from the selected location of the nuclear membrane. In addition, it is observed that the number of transcription sites is small, which is consistent with the concept that TLC1 is a strongly regulated and infrequently transcribed RNA. These spots are also more than double the average intensity of a single spot and seldom much more than three to four times more intense than the average single spot. This is consistent with small infrequent bursts. Therefore, it seems reasonable

that the two state model should apply to this gene (Figure 5.13). The burst size increases for each mutant compared to the wild-type strain. In particular the burst size increases significantly in the pop6 mutant under $30\,^{\circ}\mathrm{C}$ growth. This suggests that the pop6 mutant has a faster rate of transcription or less frequent gene inactivation It seems reasonable to assume that the mutation to pop6 had little impact on the rate of transcription of TLC1 since this effect isn't observed at the elevated temperature in the other two strains. Rather, it seems likely that the mutation to pop6 causes a change in the rate of gene inactivation. Unlike burst size the burst frequency does not seem very dependent on the change in temperature for any strain except the wild-type strain. However, the mutation does cause an increase in burst frequency in the mutants compared to the wild-type strain.

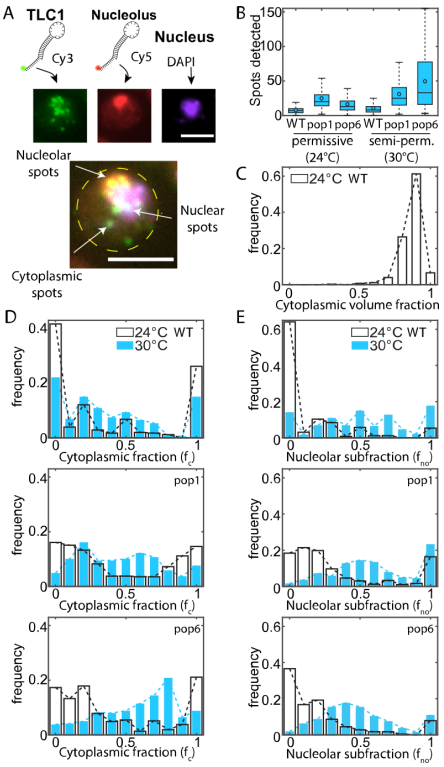


Figure 5.1: Distributions of TLC1 FISH spots. A. Schematic of the FISH method. TLC1 (green) is targeted by a short 50-nt DNA probe labeled with Cy3 at the 5' end. ITS1, a putative nucleolar marker RNA, is targeted with a 50-nt DNA oligo probe labeled with Cy5 at the 5' end. The nucleus (purple) is labeled with DAPI. A composite image of a single pop6 cell is shown from the overlap of 3 channels. TLC1 spots are assigned a localization based on the overlap with the intensity in the other two channels. The scale bar represents 5m. B. The total number of FISH spots per cell. The number of cells analyzed are from right to left 431, 485, 382, 538, 328, and 340. C. Histogram of single-cell cytoplasmic volume fraction. Uniformly localized transcripts without any spatial regulation would exhibit a similar distribution. D. Histograms of the cytoplasmic TLC1 fraction ($f_c$). The bin at zero represents the fraction of cells with no transcripts in the cytoplasm. This fraction at $f_c = 0$ is $\sim$40% for the wild-type strain but drops to $\sim$5% for the pop mutants. E. Histograms of the nucleolar TLC1 subfraction ($f_{no}$). Two-thirds of cells show no transcripts in the nucleolus at the $24\,^{\circ}\mathrm{C}$. For the mutant pop1 and pop6 strains, this fraction drops to 40% and 20%, respectively.

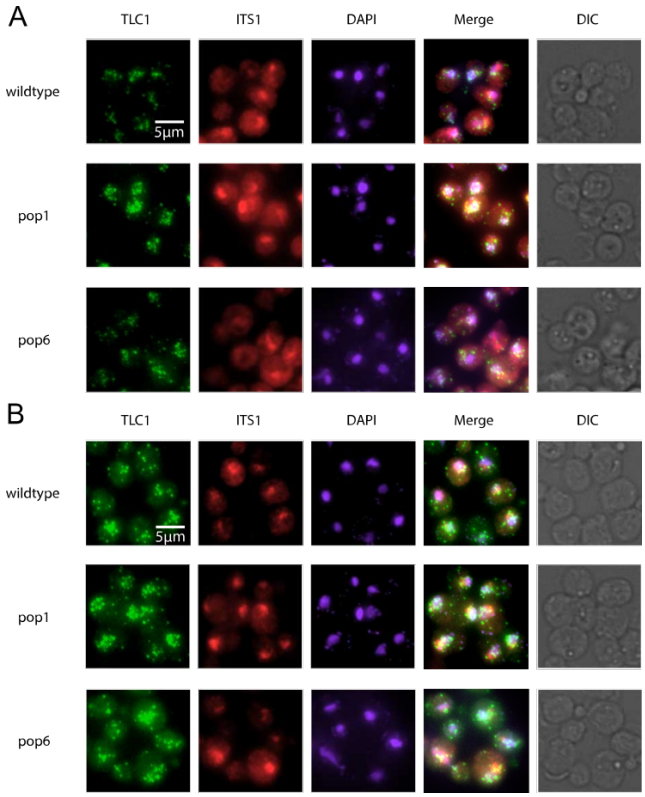


Figure 5.2: Example images of cells. A. Cells grown at $24\,^{\circ}\mathrm{C}$ (permissive growth). Images in each channel are scaled to the same contrast. B. Cells grown at $30\,^{\circ}\mathrm{C}$ (semi-permissive growth). Images in A and B are scaled to the same contrast.

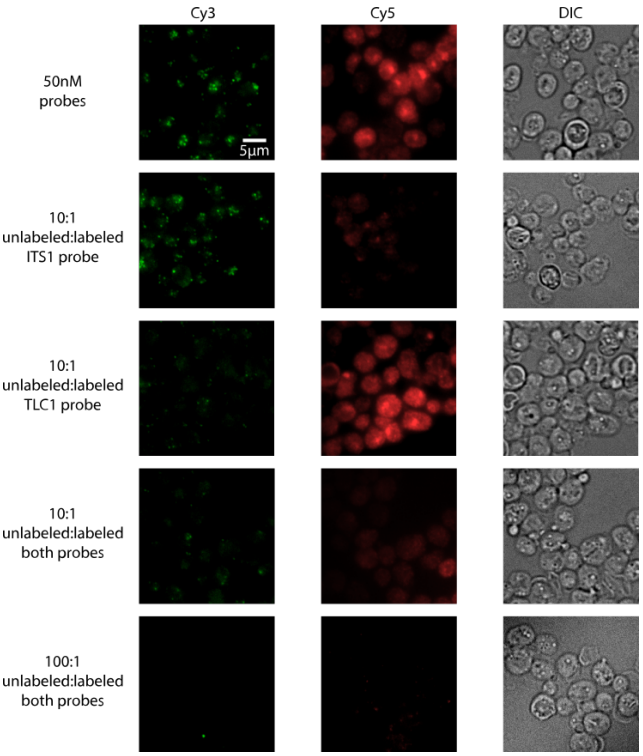


Figure 5.3: Specificity of FISH probes. Shown are example images of cells treated with several combinations of unlabeled competitor probes (probes with the same DNA sequence). All cells are from the wild-type strain grown at $30\,^{\circ}\mathrm{C}$. Cells were treated with $50\,\mathrm{nM}$ labeled probes in all cases. At 10:1 ratio of unlabeled to labeled probes, the number of spots decreases substantially. At 100:1, the number decreases even further. Each unlabeled probe affects only the FISH signal produced by its competitor probe.

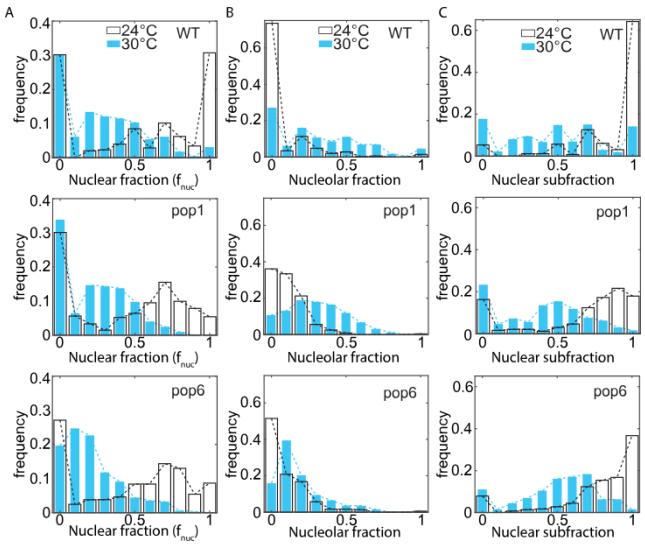


Figure 5.4: Distributions of TLC1 spot localization. A. Nuclear localization. In the wild-type strain at $24\,^{\circ}\mathrm{C}$, localization to the nucleus ($f_{nuc}$) is strongly bimodal between full ($f_{nuc=1}$) and none ($f_{nuc=0}$) with only $\sim 40\%$ of cells showing some intermediate values. The mutant strain shows an increase in intermediate nuclear localization compared to the wild-type strain. At the elevated temperature, an even higher fraction of cells show localization outside of the nucleus. B. Nucleolar localization. Wild-type strain at $24\,^{\circ}\mathrm{C}$ growth are mostly found with no transcripts in the nucleolus. While this bin remains the dominant mode of localization in the population, the mutant cells demonstrate a shift to showing some fraction of transcripts localized. At the elevated temperature all distributions including the wild-type strain demonstrate a shift towards increased localization in the nucleolus. However, there are very few instances where the cell shows full localization to the nucleolus, unlike the nucleus and cytoplasm. C. Relative decrease in the nuclear fraction. In the wild-type strain at $24\,^{\circ}\mathrm{C}$ growth, the majority of the transcripts that are within the nuclear membrane are found in the nucleus. In the mutants the distribution of cells shifts away from localizing most transcripts in the nucleus. The wild-type strain at $30\,^{\circ}\mathrm{C}$ growth show transcripts localization evenly split between the nucleus and nucleolus. The mutant strains show more localization in the nucleolus.

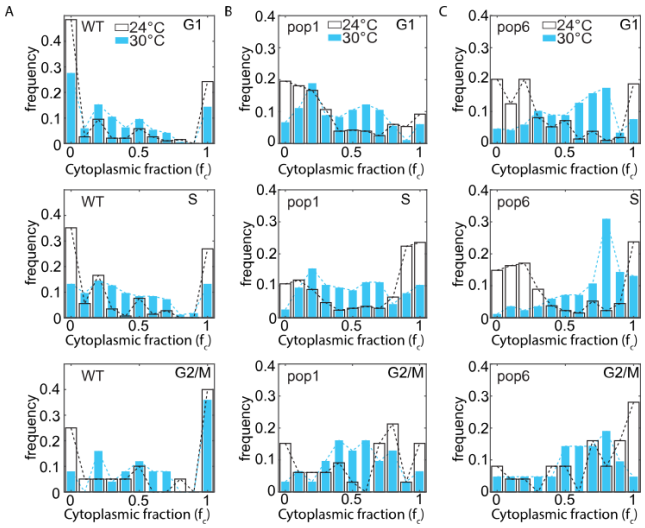


Figure 5.5: Dependence of cytoplasmic fraction ($f_c$) on the cell cycle. A. Wild-type strain. Cells were split into three categories based on cell eccentricity and nucleus size. Nuclei with eccentricity less than 0.4 were considered to be either in G1 or S phase and nuclei above that threshold were considered to be in G2/M phase. Only 10% of cells had nuclei with eccentricity above this threshold. TLC1 spots localize to the cytoplasm as the cell cycle progresses. B. Pop1 mutant. During S phase, there is an increase in localization of spots to the cytoplasm. At $30\,^{\circ}\mathrm{C}$, bimodality is suppressed at all cell-cycle phases. C. Pop6 mutant. Both mutants show an increase in localization of spots to the cytoplasm compared to the wild-type strain. However, the pop6 mutant is observed to have more spots in the cytoplasm at $30\,^{\circ}\mathrm{C}$ than the pop1 mutant.

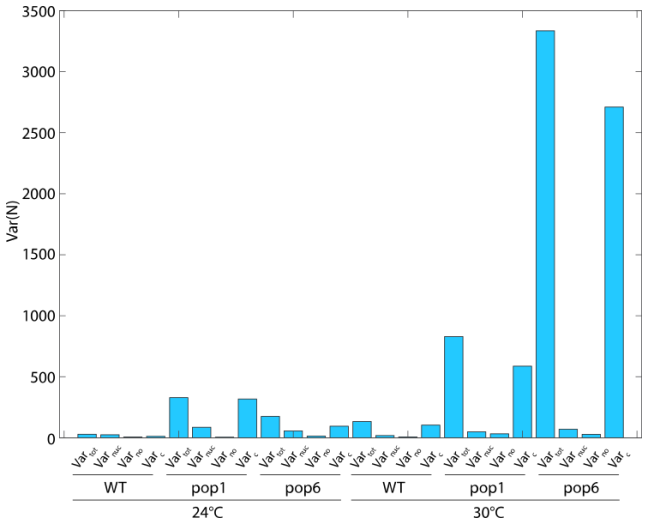


Figure 5.6: Variance of the number of FISH spots in different cellular compartments. Nuclear ($Var_{nuc}$), nucleolar ($Var_{no}$), and cytoplasmic ($Var_c$) variances are compared with each other, and the total variance ($Var_{tot}$) is shown for reference. For the wild-type grown at $24\,^{\circ}\mathrm{C}$, the nuclear variance is the largest among the three. For all other strains and growth conditions, the cytoplasmic variance is the largest

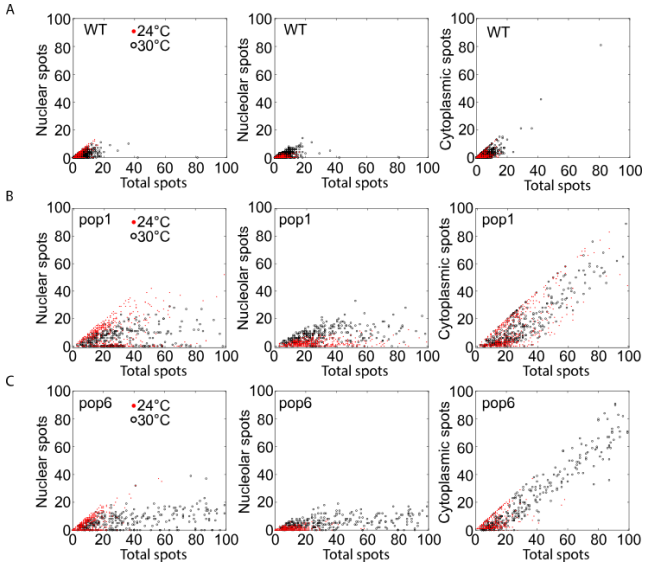


Figure 5.7: Correlation of spot counts in different subcellular compartments with the total spot count. A. Wild-type strain. For wild-type cells at permissive growth, the total spot count is strongly correlated with nuclear spot count, but weakly correlated with nucleolar or cytoplasmic spot count (red, $24\,^{\circ}\mathrm{C}$). At semi-permissive growth (black, $30\,^{\circ}\mathrm{C}$), the total spot count is less correlated with the nuclear spot count, but more correlated with the nucleolar or cytoplasmic spot count. B. Pop1 mutant. For pop1 cells at permissive growth, the total spot count is strongly correlated with the nuclear spot count with some cells showing no correlation. At semi-permissive growth, pop1 cells show reduced correlation between the total spot count and the nuclear spot count but increased correlation to the nucleolar spot count. The total spot count is strongly correlated to the cytoplasmic spot count. C. Pop6 mutant. For pop6 cells at permissive growth, the total spot count is increased compared to the wild type but has the same pattern of correlation. At semi-permissive growth for pop6 cells, the total spot count shows no correlation to either nuclear or nucleolar spot counts but is strongly correlated with the cytoplasmic spot count.

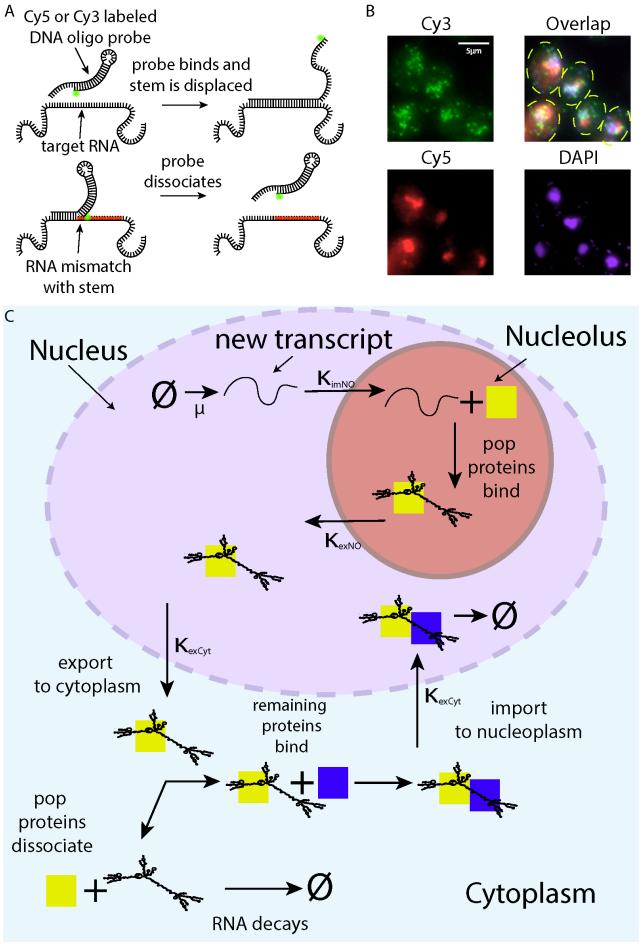


Figure 5.8: Schematic of TLC1 lifecycle. A. TLC1 probe schematic. TLC1 is targeted using a hairpin probe to increase specificity. B. Example cells. Cells are shown from the pop6 mutant for each of the three channels used in this experiment, Cy3, Cy5, and DAPI. C. Lifecycle of TLC1. A transcript is generated in the nucleus and must enter the nucleolus to bind pop proteins. The transcript is then exported to the cytoplasm to bind additional proteins. Once the mature telomerase enzyme is formed, the transcript is imported back to the nucleus to perform its function. At each point, there is the possibility of protein dissociation and RNA degradation.

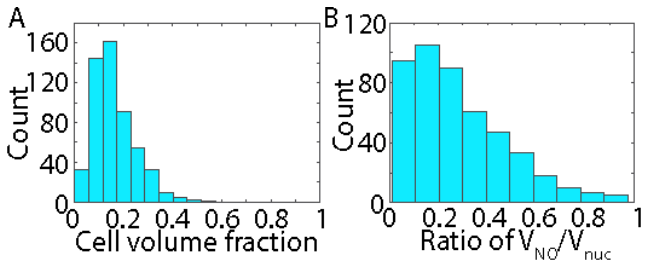


Figure 5.9: Relative size of Nucleus and Nucleolus. A. Nucleus size. The size of the intensity region selected as the nucleus based on the DAPI image is shown as a histogram of volume fractions. The fraction was calculated using the segmented region of the cell. B. Nucleolus volume. The the most intense region of the cell in the CY5 channel that was also near the nucleus is compared to the size of the nucleus selected in the DAPI channel. The mean ratio is 28.67±0.020.

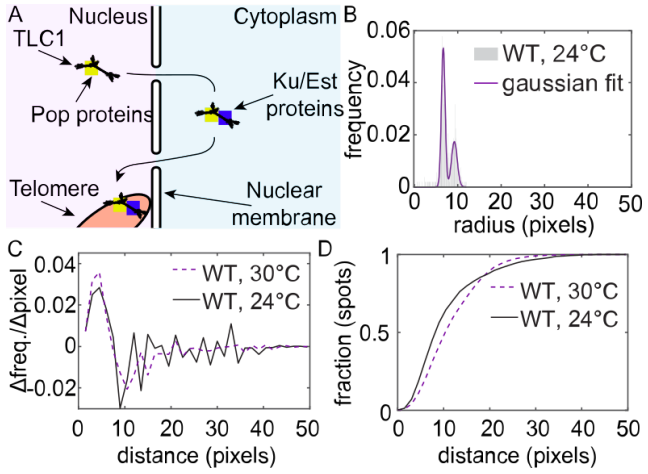


Figure 5.10: Choosing the appropriate nuclear boundary. A. TLC1 activity near the nuclear membrane. The TLC1 transcript has several factors in its lifecycle that drive it to the nuclear membrane. The need for Ku/Est proteins drives it to the membrane when it is in the nucleus. The need to maintain the telomeres drives it to the membrane when it is in the cytoplasm. Finally, the mature enzyme localizes to the telomeres which are attached to the nuclear membrane when it returns to the nucleus. B. Histogram of nuclear radii. The radius of the nucleus for wild-type cells grown at 24 °C is shown as a histogram and fit with a sum of two Gaussians. The radius is between seven and ten pixels. This equates to a nucleus that represents ∼15% of the cellular volume. C. 2nd derivative of the cumulative radial distribution. To illustrate the correspondence of spot location and the radius selected, the second derivative of the cumulative radial distribution of spots is shown. There is a peak corresponding to the first peak of selected nuclear radii and a local minimum at the second peak. D. Cumulative radial distribution. The cumulative distribution of spots shows that there is a difference in relative diffusion with regards to the nuclear membrane for cells grown at 30 °C

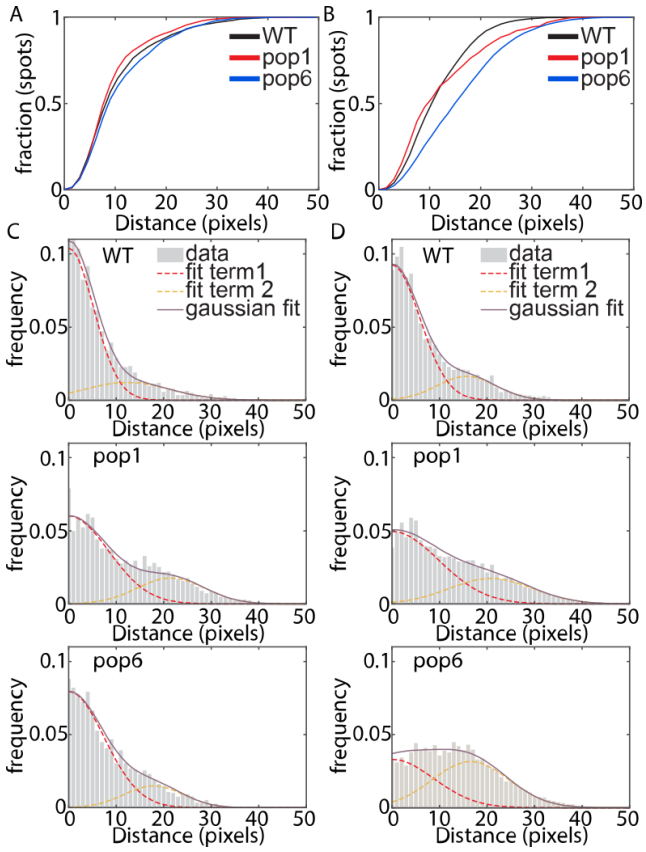


Figure 5.11: Diffusion of TLC1. A. Cumulative radial distribution at $24\,^{\circ}\mathrm{C}$. The detected spots in the entire experiment are represented along the axis of the minimum distance to the nuclear membrane selected in the DAPI channel. B. Cumulative radial distribution at $30\,^{\circ}\mathrm{C}$. Both the pop1 and pop6 mutant are distributed further from the nuclea membrane than at the $24\,^{\circ}\mathrm{C}$ condition and compared to the wild-type cells. C. $24\,^{\circ}\mathrm{C}$ sum of Gaussians. All three strains are fit with a sum of Gaussians, which has a very small $\chi^2$ in all cases.

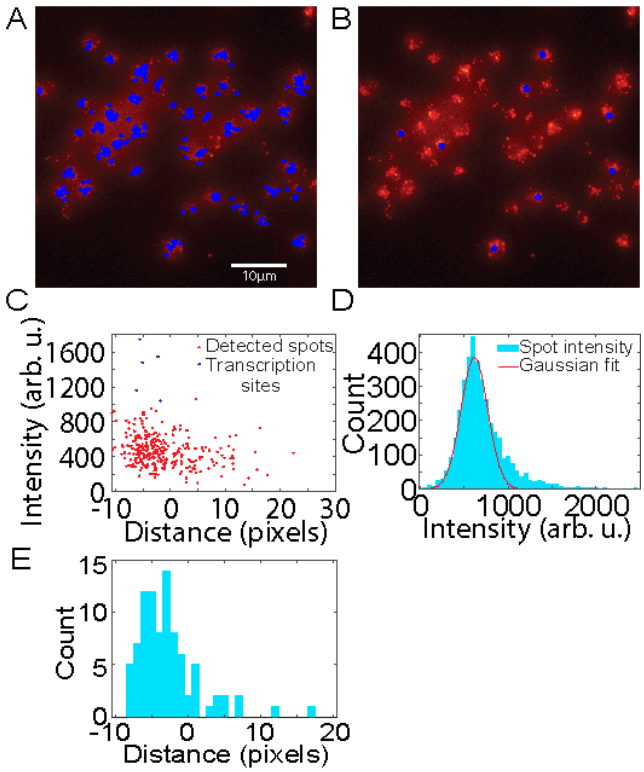


Figure 5.12: Transcription site selection. A. Spots detected. All spots in the representative image are shown with a blue mark. B. Transcription sites. The transcription sites are shown in blue. There is only one site per cell and every site is within the nucleus by at least 2 pixels from the selected border. C. Intensity versus distance from nuclear membrane. The distance is calculated by fitting spots with a Gaussian profile and extracting sub pixel coordinates. The intensity of each spot is shown with regards to the distance away from the nuclear membrane. Transcription sites are designated with blue. D. Intensity of spots. The intensity of every spot detected in the experiment is shown with background subtracted. The distribution is unimodal with a small fraction of spots showing intensities that are larger than expected from Gaussian distributed intensities. E. Location of spots that are brighter. The spots in the tail are selected at intensity values over 1000. The majority of these spots are at the edge or within the nucleus.

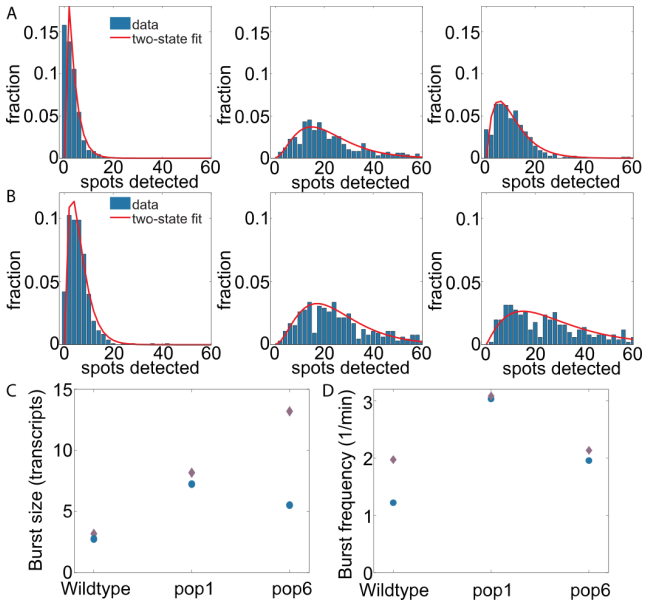


Figure 5.13: Two state model of TLC1 transcription. A. Fit of 24 °C cells. The fit of wild-type (left), pop1 mutant (middle), or pop6 mutant (right) cells is shown in red. B. FIt of 30 °C cells. The strains show increased number of transcripts in all cases. C. Burst size. The number of transcripts produced during each transcription event defined as the ratio of the rate of active transcription to the rate of gene inactivation. The burst size increases in the mutant strains and at increased temperature with a more dramatic temperature based change in the pop6 mutant. D. Burst frequency. The burst frequency is defined as the ratio of the rate of gene activation to RNA decay. The burst frequency increases in both mutants and at increased temperature.

# CHAPTER 6

# STRAND DISPLACEMENT BASED BARCODING OF RNA ISOFORMS

Alternate splicing variation can lead to families of transcripts that share significant sequence homology, but are not separably observable through conventional FISH methods. An interesting class of non-coding transcripts that is of broad interest due to its potential impact as a biomarker of cancer[82] is circular RNA, which are either covalently closed circularized exons or intron lariats which are closed by a different mechanism[83]. Most of the understanding of circular RNA and its ubiquity in cells across all eukaryotes comes from ensemble methods such as bioinformatics, PCR based methods, or gel electrophoresis. The biological function of these transcripts is poorly understood in general, but the demonstration that CDR1AS functions as a miRNA sponge suggests that circular RNA can perform key regulatory functions. Broader understanding of these transcripts and their biology requires the ability to overcome the technical limitations of detecting them.

RNA FISH is a robust method that has been used for several decades to quantify the distribution of RNA transcripts. This method involves targeting an RNA transcript with a multiplicity of fluorescently labeled DNA oligonucleotide probes in order to observe a strong punctate signal. The drawback of this method is that it requires at minimum $\tilde{2}00$ nucleotide sequences to detect RNA with greater than 98% efficiency. When considering alternate splicing variations, conventional FISH methods are further limited by the lack of long exons that are unique to the isoforms. In principal, spatial barcoding could enable the detection of isoforms with adequate length of targets, but these methods also suffer from increased acquisition and analysis requirements. The joining of the 5' and 3' ends of the RNA is the most unique feature to these transcripts over their linear homologues. One approach targeting these junctions would be to design a probe that to single nucleotide specificity is to use a probe masked to expose only a short ten nucleotide toe-hold in conjunction with

a set of guide probes to determine false positives. In this case, there is no barrier to strand displacement and the assumption is made that the rate of strand displacement is slower than the rate of dissociation of toe-holds with single mismatches. If instead the site of interest is placed at the first base pair in the masked region, a mismatch will massively slow down displacement kinetics and multiple mismatches with prevent displacement completely.

We demonstrate using the SUS1 gene in *saccharomyces cerevisiae* that the previous limitations of conventional RNA FISH methods can be overcome to detect alternate splicing variations and circular RNA. We use DNA strand displacement as a mechanism for specific detection of RNA isoforms even in cases of short sequences and high homology between isoforms. We further extend this method using a universal probe and a secondary displacement interaction to sequentially barcode the NURF1 gene in *caenorhabditis elegans* an organism that exhibits alternate splicing in a large number of genes. Additionally, the splicing variants in *C. elegans* are expressed in normal growth conditions and with some spatial dependence on location in the germline of *C. elegans*. The work with *C. Elegans* is done in collaboration with Patrick MacGrath and his student Wen Xu.

## 6.1 HCR on yEVenus

One of the downsides of the sFISH method was the inability to achieve high detection efficiency. A way to deal with this is signal amplification. This would need to be in a way that ensured that there was an actively fluorescing dye molecule on each target since we are not limited by our detection sensitivity. Hybridization Chain Reaction (HCR) is a means to achieve enzymeless amplification of a nucleic acid sequence using strand displacement reactions of a pair of meta-stabel probes. The performance of HCR was compared to the performance of sFISH and mpFISH. In this scheme, the binding site chosen for sFISH is used as the genomic portion of the HCR probe. The stem of this probe is chosen to be 16 nucleotides based on the results from Choi, et al[84]. They determine that the HCR probe set is meta-stable when the ratio of stem to hairpin loop length is 1.6 or greater.

We determine using the high and low expression strains of yEVenus that the HCR detection scheme achieves the same spot count as mpFISH at a working concentration of approximately $50\,\mathrm{nM}$, which is also where the number of spots saturates (Figure 6.2). Another interesting observation is that using mpFISH at concentrations over $100\,\mathrm{nM}$ tended to produce an extremely high background, whereas, in the case of HCR it produces extremely bright punctate spots. This might be an effect of having binding sites always be active. The number of photobleaching steps also increases over the concentration range (Figure 6.3).

One question with regards to HCR is how does a mismatch in the stem with respect to the target RNA effect the ability of HCR probes to amplify. To determine this we designed two additional pairs of HCR probes which had a mismatch with respect to the target at either the first one or two basepairs of the stem of the hairpin probe. The hairpin probes were not mismatched with themselves. This mismatch produced very few amplified spots for a single mismatch and no amplified spots for two mismatches (Figure 6.1D).

## 6.2 SUS1 isoforms

Since HCR probes are capable of specifically detecting mismatches in the stem and therefore not proceeding to amplify when the sequence is mismatched, this allows to design a probe that can detect if two pairs of sequences are adjacent. An interesting example where this is the only unique identifier is circular RNA. There are only 11 genes with two exons which can produce circular transcripts of the middle exon during splicing in the right circumstances. In yeast there are no examples of circular transcripts to my knowledge that are expressed through any other mechanism. It has been shown through RNase resistance that circularization of the majority of these two intron genes is possible with SUS1 demonstrating the highest relative expression[85]. A set of HCR probes is designed based on the SUS1 sequence to target the junctions between each exon as well as the 5' and 3' ends of the middle exon (Figure 6.4A). The aneuploid mutant strains with one to four copies of the genome as well as the yEVenus strains were both checked for circular RNA expression and

Table 6.1: SUS1 probe sequences. HCR prbes sequences are shown for the three binding sites on SUS1 isoforms.

| Melting Temp($^{\circ}C$) | Sequence | |
|---|---|---|
| Circ HCR1 | 47.7 | [Alexa488]GTTTGAAATTCTAATGCTTTGGGGTACTCATACCAAAGCATTAG |
| Circ HCR2 | 49.7 | CCAAAGCATTAGAATTTCAAACCTAATGCTTTGGTATGAGTACC |
| J 1-2 HCR1 | 51.1 | [CY3]GTTTGAAATTAGTTCATAGTTTTGGACTAGCAAAACTATGAACT] |
| J 1-2 HCR2 | 47.8 | AAACTATGAACTAATTTCAAACAGTTCATAGTTTTGCTAGTCCA |
| J 2-3 HCR1 | 51.1 | [CY5]CTAATGCTTTGGGATACCATTTCATAGTGACCAAATGGTATCCC |
| J 2-3 HCR2 | 47.8 | AAATGGTATCCCAAAGCATTAGGGGATACCATTTGGTCACTATG |

intron retention by incubating cells that had just reached 0.6 $OD_{600}$ for up to two hours at $37\,^{\circ}\mathrm{C}$. After incubating the cells for just $30\,\mathrm{min}$, the number of SUS1 mature transcripts had reduced and the number of circular transcripts and intron retentions had increased. The expression of SUS1 increases monotonically with the number of copies of the genome. The aneuploidy mutants express more intron retention and circular isoforms at each time point as well as saturate in circular RNA expression faster. When the Cy3 and Cy5 signal are colocalized, the transcript is classified as a mature transcript that has had both introns spliced out of it (Figure 6.5). When incubated at the higher temperature, the amount of colocalization changes from being mostly spliced transcripts to being mostly misspliced transcripts. Even at $90\,\mathrm{min}$ of heat shock, the cells still demonstrate some colocalization of Cy3 and Cy5 spots. However, there are also some spots that colocalize either Alexa488 with Cy3 or Cy5. In these cases, it is unclear what has happened. Most likely there is a circular transcript near a misspliced transcript, however, further experiments would need to be done to verify this.

## 6.3 Sequential detection of NURF1 isoforms

### 6.3.1 Attaching *C. elegans* to the coverslip

Some labs use Superfrost+ slides to adhere *C. elegans* for cryo-EM experiments. These slides come with a strong positive charge that is created by a proprietary silanization method. When worms are placed on these slides they become strongly adhered and can be handled over extended periods with little regard to the adherence of the worms. When

the worms are placed on the slide, the worms are dissected by cutting at the head which causes the intestine and germline to eject due to pressure. Worms are kept in methanol before and during dissection and slides are evaporated prior to handling. The downside of this method of adherence is that the worms are approximately $20\,\mu\mathrm{m}$ thick and have significant variation in height from the distal end to the oocytes in the germline.

To place the worms at the surface closest to the the objective, functionalization of the coverslip is necessary. Other options tried included drying the worms on a plasma-etched surface, embedding in a thin layer of $2\%$ agarose, trapping under a cellulose membrane from a dialysis tube, and using poly-L lysine coated coverslips. The only option in that list that worked well was poly-L lysine, which was prepared using a protocol from Wormbase using high molecular weight poly-L lysine and sodium azide to functionalize the surface. This method of attachment does not last long enough to perform many hybridizations over a period of two weeks. A final method was to functionalize the surface with (3-Aminopropyl)triethoxysilane (3-APTES). The protocol for this is as follows.

1. Wash the slides in anionic detergent for $10\,\mathrm{min}$ and rinse with DI water. Alternatively, etch in plasma etcher for $10\,\mathrm{min}$.
2. Sonicate in ethanol for $15\,\mathrm{min}$.
3. Sonicate in DI for $15\,\mathrm{min}$.
4. Sonicate in $1\,\textsc{m}$ NaOH for $30\,\mathrm{min}$.
5. Rinse with DI.
6. Dry using vacuum pump in plasma etcher chamber.
7. Place coverslips in a glass petri dish at angle using a clean glass slide.
8. Dry completely by placing dish on a hot plate at $170\,^{\circ}\mathrm{C}$ for $1\,\mathrm{h}$.
9. Place dried coverslips into a coplin jar that is completely dry.

10. Fill coplin jar to the top with $\sim 50\,\mathrm{mL}$ ultra dry Toluene.

11. Add $1\,\mathrm{mL}$ to $5\,\mathrm{mL}$ of 3-APTES and cover.

12. Incubate at room temperature for $1\,\mathrm{h}$.

13. Dispose of Toluene and rinse with Toluene. (dispose in organic waste)

14. Rinse with ethanol.

15. Rinse for $1\,\mathrm{min}$ with DI.

At this point, coverslips are stored in a vacuum and used within several hours. Attempts to keep coverslips for longer have failed and it is recommended to make them immediately before use.

A further consideration is the choice of tape. The removable double stick tape (3M) will detach after several hours of exposure to acqeuous solution. It was necessary to use a chemical resistant tape (3M, catalog number: 9671LE) that will last indefinitely under the conditions for FISH. This tape and attachment has been demonstrated to last for several weeks and at least fourteen hybridizations (Figure 6.8A).

### 6.3.2 NURF1 probe design

The original design for NURF1 probes was to target individual exons and use HCR as just an amplification method. The desire was to label three exons with spectrally disctinct fluorophores and then see the colocalization of these probes to identify the NURF1 isoform. These exons were chosen so that the middle probe (Cy3 label) was shared between all isoforms and the identity would be determined by its colocalization with either end labeled with Alexa488 and Cy5. This was unsuccessful. The only probe that could consistently be detected was Cy5. This is due to the lower background fluorescence in the red spectrum for this organism. The failure of this initial design could be due to either the background fluorescence or the poor quality of amplification. mFOLD predicted a stable hairpin form

of the probe with a stem that was twelve nucleotides long. This could have had some impact on the ability of the HCR probe set to amplify. If they were not metastable, it would be expected to see many intense spots throughout the sample. However, this did not seem to be the case. In either case, the quality of HCR amplification cannot be ruled out as a problem without further experiments.

### 6.3.3 Sequential hybridization

Since the spectral barcoding approach was unsuccessful a redesign of the HCR detection scheme for NURF1 was performed. In this new scheme, cost would be reduced by designing a universal probe labeled with Cy5 and aligned to both the yeast and *C. elegans* genomes to have minimal overlap for a 30 nucleotide sequence (Figure 6.6). A binding site for this probe and for an additional oligo to act to initiate deconstruction of the HCR polymer is included at one end of the first HCR probe (Figure 6.7). Therefore, any set of HCR probes could be ordered without a fluorophore, which is the most expensive part of oligo synthesis.

FISH is performed by flowing approximately $20\,\mu\mathrm{L}$ to $40\,\mu\mathrm{L}$ of hybridization buffer with a working concentration of Cy5 probe and HCR probes of $50\,\mathrm{nM}$. This seems to be the concentration where the background is lowest and number of amplified spots is highest. Currently, concentrations of 5,10,20,50 and $100\,\mu\mathrm{M}$ have been tested with further analysis needing to be performed. The goal is to find the concentration where the number of spots appearing in subsequent hybridizations with the same probe are maximized. Subsequent hybridization is performed by flowing imaging buffer containing a working concentration of $1\,\mu\mathrm{M}$ displacer probes into the chamber and incubating at room temperature for $15\,\mathrm{min}$ (Figure 6.8B). This probe binds to the end of the initiator, which has a six nucleotide toehold (Figure 6.7B). This toehold is purposefully chosen to be short and unstable so that any displacers that do not get washed away do not impact subsequent hybridizations. Once the HCR amplified polymer is deconstructed, the products are unbound Cy5 labeled

probes, unbound amplifier probes, and a fully doublestranded construct of the initiator and displacer. In principal, any unwashed amplifier should not be able to create false positives since it is much more likely to form a hairpin again then to bind to anything else in a subsequent hybridization. We demonstrate that the signal from each hybridization can be fully reset by a wash with stringent (15% formamide) wash buffer and incubation at $37\,^{\circ}\mathrm{C}$ to allow time for unbound probes to diffuse out of the sample. It appears that the distribution of the FISH spots for each exon is different (Figure 6.8C). Further analysis methods are being developed to colocalize transcripts across a multi-region image using an affine transformation to move spots located in one acquisition into the same reference frame as the other acquisitions.

### 6.3.4 Binding of HCR probes under extremely high expression

The HCR probe amplification represents an infinitely on binding site that alternates between two states. If the supply of the two probes is not a limiting factor, then during the entire hybridization incubation every site is always available to one of the two probes. This in general presents very little difficulty since the number of potential targets to bind for most genes is small (¡100). However, an interesting case is found when considering HCR for a gene like ITS1, where there are thousands of copies spliced out per minute since there is one copy spliced out of the rRNA per ribosome created. In addition the nucleolus is disassembled during mitosis so that the transcripts while mostly concentrated at the nucleolus should be present as a pretty significant background throughout the cell. However, this was not observed when attempting HCR on ITS1. In addition, establishing a working concentration was quite difficult since at low concentrations there was still a significant amount of probes throughout the cell and a high concentration created a background that was too bright for the EMCCD camera let alone allowing for any contrast with the nucleolus. This was true when only the initiator probe was used as well.

To illustrate one possible explanation a simulation of probe biding on a 60x60 grid rep-

resenting an image the size of a cell is performed through a monte carlo method combined with a random walk on a 2D lattice. This shows that with a strong binding probability in the middle of the square and weak binding probability every where else it should be fairly easy to gain contrast between the two binding sites unless the weak binding probability is significant. In this case a depletion region appears in the interior of the cell (Figure 6.9A-B). These cases do not recapitulate the observation. However, there is no reason to assume that the binding sites should be perfectly homogeneous throughout the exterior region. If the binding sites are randomly distributed but still significant compared to the total number of binding sites then there is a case where the binding to the weak sites and strong sites provides little contrast (Figure 6.9).

Table 6.2: NURF1 probe sequences. Probe sequences for the universal Cy5 probe and HCR and deconstructor probes for the junction between the 2nd and 3rd exons, 21st and 22nd exons, and the 24th and 25th exons. These probes allow for the unique identification of 5 isoforms the full length isoform (1,1,1), mid length isoforms (1,1,0) and (0,1,1), as well as short isoforms (1,0,0) and (0,0,1) which have all been annotated on Wormbase.

| Identity | Melting Temp($^{\circ}C$) | Sequence |
|---|---|---|
| Universal Cy5 | 65.2 | /5Cy5/ATTACACGGTAGGCCCTGGCCCTAGATTGA/ |
| J. 2-3 Initiator | 72.5 | ACTCGTCGTCCTTCTACCCAATCGCAGAAGGACGACGAGTTTTGCATCATTCAATCTAGGGCCAGGGCCTACCGTGTAATGTGAGT |
| J. 2-3 Amplifier | 51.1 | GCGATTGGGTAGAAGGACGACGAGTATGATGCAAAACTCGTCGTCCTTCT |
| J. 2-3 Deconstructor | 47.8 | ACTCACATTACACGGTAGGCCCTGGCCCTAGATTGAATGAGTCAAAACTCGTCGTCCTTCTGCGATTGGGTAGAAGGACGACGAGT |
| J. 21-22 Initiator | 49.2 | CGGCCAAACCAAAAGACCCAATCGCCTTTTGGTTTGGCCGCCCGATGCGGTCAATCTAGGGCCAGGGCCTACCGTGTAATGTGAGT |
| J. 21-22 Amplifier | 47.8 | GCGATTGGGTCTTTTGGTTTGGCCGCCGCATCGGGCGGCCAAACCAAAAG |
| J. 21-22 Deconstructor | 49.2 | ACTCACATTACACGGTAGGCCCTGGCCCTAGATTGACCGCATCGGGCGGCCAAACCAAAAGGCGATTGGGTCTTTTGGTTTGGCCG |
| J. 24-25 Initiator | 49.2 | CCAACGGAATCAATCACCCAATCGCGATTGATTCCGTTGGCTTCTGTTCCTCAATCTAGGGCCAGGGCCTACCGTGTAATGTGAGT |
| J. 24-25 Amplifier | 47.8 | GCGATTGGGTGATTGATTCCGTTGGGGAACAGAAGCCAACGGAATCAACGGAATCAATC |
| J. 24-25 Deconstructor | 49.2 | ACTCACATTACACGGTAGGCCCTGGCCCTAGATTGAGGAACAGAAGCCAACGGAATCAATCGCGATTGGGTGATTGATTCCGTTGG |

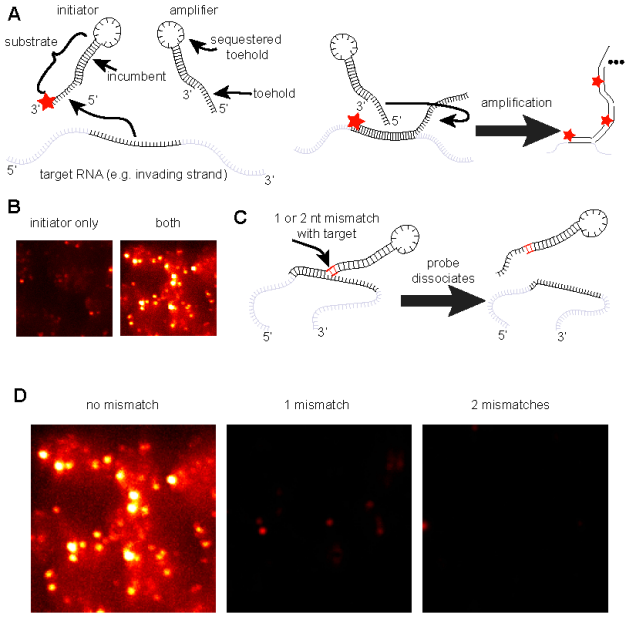


Figure 6.1: Hybridization Chain Reaction (HCR) design and specificity. A. A hairpin probe is designed to be metastable with another complementary hairpin probe in the hybridization condition. The hairpin probe consists of a short 10nt toehold and 16nt stem with an additional toehold sequestered in a loop. When the input RNA is present the initiator probe opens and the amplifier probe can then bind. B. Comparison of hairpin probe with and without amplification. When the amplifier probe is not present, the HCR reaction cannot proceed and the sample displays only single fluorophores. When the amplifier is present, the signal is greatly increased and the majority of spots are due to multiple fluorophores. C. Schematic of input mismatch. The HCR system is designed so that the initiator and amplifier are fully matched. Either one or two mismatches with the input RNA are present at the leading basepair of the stem. D. Mismatch dependence of amplification. With no mismatch, amplification proceeds successfully. When a with one or two mismatches amplification does not proceed.

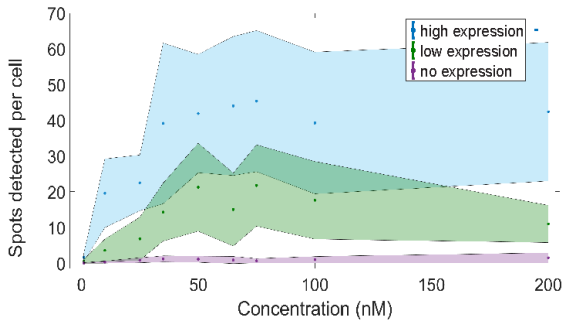


Figure 6.2: Spot dependence on HCR probe concentration. The control for yEVenus is compared to the low and high expression strains over a range of working concentrations. The number of spots detected at saturation is consistent with the multiple probe experiment.

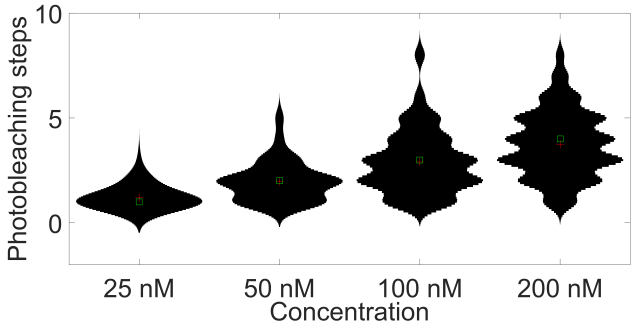


Figure 6.3: Photobleaching steps of yEVenus HCR probes. A concentration series was performed to determine the effect of concentration on HCR. The number of photobleaching steps increases from an average of 1 to an average of 4 from $25\,\mathrm{nM}$ to $200\,\mathrm{nM}$

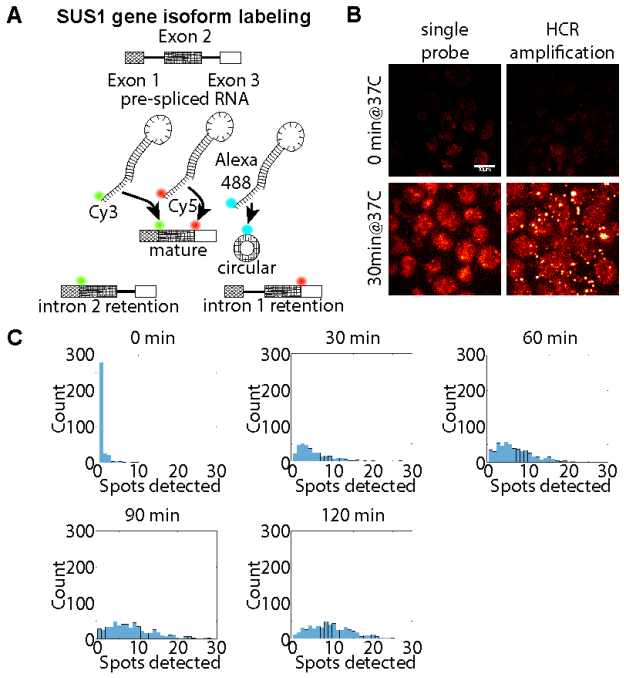


Figure 6.4: Direct detection of circular RNA. A. Probe schematic. The SUS1 gene contains two introns. Probes are designed to target the junctions between the first and second exon, the second and third exon, and the 5' and 3' ends of the middle exon. B. Temperature and amplification dependence of signal. Cicular RNA are targeted with Alexa488, which is very dim compared to Cy3 and Cy5. Alexa488 has a molar extinction coefficient less than half of either of the other fluorophores (73,000). When cells reach 0.6 $OD_{600}$, they are then switched from $30\,^{\circ}\mathrm{C}$ to $37\,^{\circ}\mathrm{C}$ for up to $120\,\mathrm{min}$. After $30\,\mathrm{min}$, the cells demonstrate significant accumulation of HCR dependent signal, which represents the accumulation of circular RNA. C. Time dependence of signal. The accumulation of circularized SUS1 trancripts increases during the first $90\,\mathrm{min}$ where it seems to plateau.

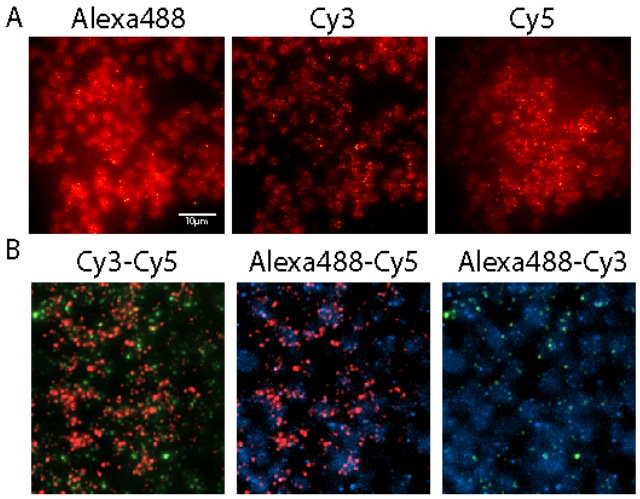


Figure 6.5: Colocalization of FISH spots for circular RNA. A. Individual channels. Alexa488, Cy3, and Cy5 were excited in succession at $100\,\mathrm{ms}$ exposure times. B. Colocalization of spots after 90 minutes of heat shock. Overlap images are shown of the central region of intensity for Alexa488, Cy3, and Cy5. After the cells have been incubated at $37\,^{\circ}\mathrm{C}$ for $30\,\mathrm{min}$, the majority of Cy3 and Cy5 are not colocalized. Alexa488 spots are occasionally colocalized with either Cy3 or Cy5 spots.

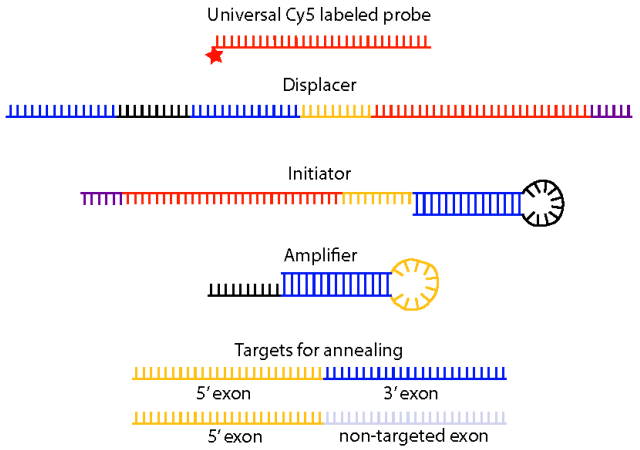


Figure 6.6: Schematic of strand displacement probes and targets. Red designates sequences that are cognate to the Cy5 labeled probe. The toehold to initiate displacement is purple. Black is a randomly generated sequence to construct the HCR probe. The genomic sequence is colored yellow for the 5' exon and blue for the 3' exon. Non-specific sequences are designated in grey.

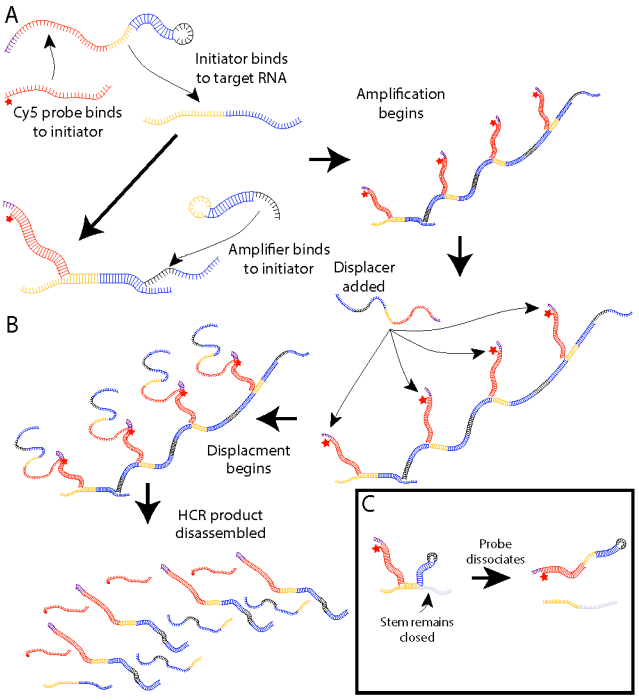


Figure 6.7: Schematic of the reactions for sequential RNA isoform detection. A. HCR reaction. The HCR probes and universal Cy5 labeled probe are added to hybridization buffer and flowed into the chamber and incubated overnight. B. Post-acquisition. The displacer is mixed with imaging buffer and flowed into the chamber at 1 µM working concentration. C. Non-specific targets. The HCR system is metastable in the hybridization condition and can only weakly interact non-specifically. The length of the stem determines the effect of any interaction with the sequestered toehold. The 30nt Cy5 labeled probe is aligned with the genome of yeast and *C. elegans* to have minimal overlap.

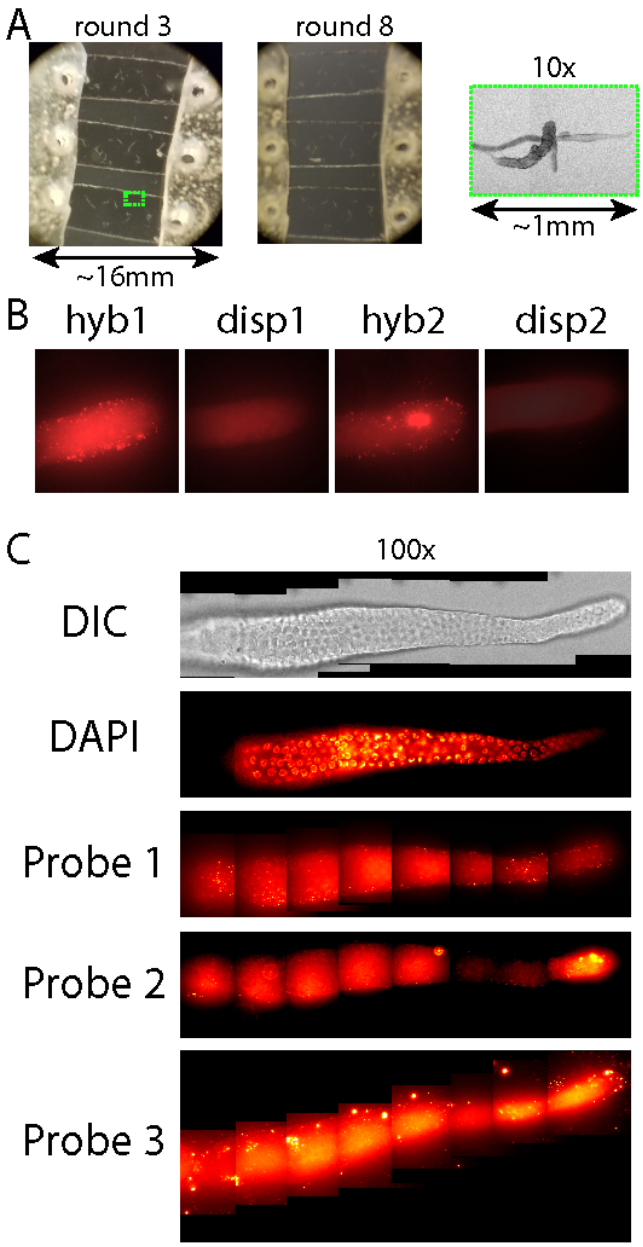


Figure 6.8: Sequential hybridization of NURF1 isoforms. A. Worm adherence. Two images are shown from taken with a cell phone and 10x ocular. Worms stay adhered under $1\,\mathrm{mL}$ buffer exchange via pipette. A DIC image acquired with the 10x objective is shown with green border showing the location of the worm in the larger image. B. Sequential hybridization and displacement. The same section of the worm is acquired multiple times. Images of the worm after displacement are taken after $15\,\mathrm{min}$ incubation with $1\,\mu\mathrm{M}$ displacer and a wash followed by $30\,\mathrm{min}$ incubation at $37\,^{\circ}\mathrm{C}$. Incubation after a wash is not adequate to remove the signal. C. Hybridization with three probes. A universal Cy5 probe is hybridized to 3 different sets of HCR probes. The initiator contains the sequence for the labeled probe. Each image is a constructed from eight fields of view that significantly overlap with each neighboring frame.

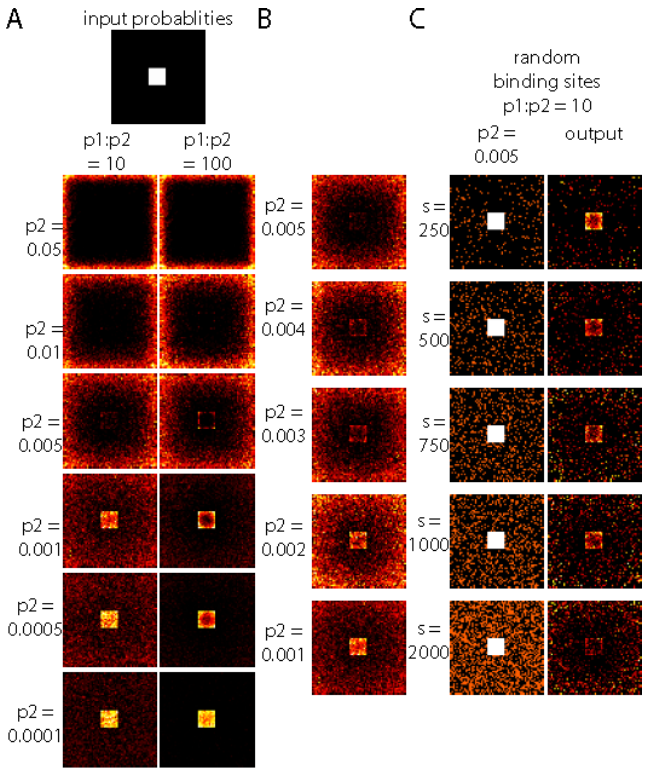


Figure 6.9: Hybrid Gillispie algorithm. A. Representative trials. The input image representing the binding probabilities is shown (top). There are 10000 events which are determined by a monte carlo simulation of binding and time to bind. Here the ratio between the inner and outer pixel probability is either 10 or 100. When the probability of binding is high, the trajectories are short and binding events occur mostly in the outer edge. When the probability is low the trajectories are long and they tend to end at the higher probability pixels. B. Intermediate probabilities. Further representative simulations between the middle points of (A) for the p1:p2 = 10 case. There is a point where the high probability and low probability pixels have similar amounts of binding except for a region of depletion around the inner square. C. Randomly distributed outer binding sites. Rather than setting the probability of each pixel to be the same. The sites for the outer region are randomly distributed. The number of binding sites ranges from 350/3600 to 2100/3600. The effect of depletion around the inner region is reduced when the number of binding sites is significant compared to the total number of sites.

# CHAPTER 7
# CONCLUSION

Single molecule FISH is a robust method that is unsuited to the characterization of short or highly homologous transcripts. In this work, we have demonstrated that a single fluorophore is capable of detecting $\sim$60% of transcripts. Further, we demonstrate that this method is well suited to common single molecule tools such as FRET. Using smFRET we were able to establish the detection efficiency of a single probe and directly demonstrate that the FISH signal is due to probe binding without needing to use intensity thresholding. We also demonstrate that this method can provide significant information about the life-cycle of the non-coding transcript TLC1 without needing to have the absolute transcript count. This work represents the first demonstration of FISH with a single fluorophore. We also demonstrate Hybridization Chain Reaction (HCR) amplification in both yeast and *C. Elegans*. This technique is then leveraged to exploit strand displacement kinetics and has shown single nucleotide specificity at the leading basepair in the stem of the HCR probe. To demonstrate the importance of this specificity, the RNA isoforms of SUS1 in yeast and NURF1 in *C. Elegans* are characterized. Furthermore, the addition of a secondary binding site and a means to initiate an additional strand displacement reaction demonstrates a cheap and efficient means to fully characterize the isoforms of any RNA transcript.

By sacrificing the robustness of transcript detection, we were left with the necessity to naively accept every peak in fluorescence intensity to be a transcript. This meant that it was critical to understand how to accept this with confidence required careful consideration of each step of the FISH protocol. It was then necessary to address the non-specific spots found in a negative control. From this we determined that formaldehyde treated cells interact non-specifically with FISH probes at a higher rate with comparison to a methanol fixation method. We also determined that the spectral purity of reagents in either formaldehyde

or methanol fixation was significant to removing false positives. This constraint makes this method much simpler than most FISH methods since they often require extensive testing of probe sequences, fixative conditions, and hybridization conditions to achieve a good signal. Another significant benefit is the ability to use a single cognate unlabeled competitor probe as a control for false positives. This would be unfeasible for a conventional FISH protocol due to needing up to 48 cognate probes for competitors thereby significantly increasing cost.

When applying this method to the non-coding transcript TLC1, it was necessary to consider the integrity and size of the fixed cells. Methanol fixation decreases cell volume making imaging more difficult. It also has some impact on the nuclear membrane which may degrade its integrity. Since location relative to the nucleus and nucleolus was desired and transcripts were likely to be concentrated to a sub-cellular localization formaldehyde fixation was revisited. We observe that the use of higher quality formaldehyde and a hairpin probe yield $\sim$1 non-specific transcript per cell. This was important because having a low rate of false positives enables this method to make inferences based on the location of the detected spots with regards to either organelle. We observe that under mutation and increased temperature, the localization of TLC1 transcripts shifts from being mostly in the nucleus to being mostly in the cytoplasm. We also observe that the distance of a transcript in the cytoplasm relative to the nuclear membrane increases compared to the wild-type strain. In addition, the two state model indicates that the burst size and burst frequency increase due to the mutation.

HCR was used on several different transcripts and strains including yEVenus and SUS1. The initial goal for the use of amplification was to determine whether or not a single binding site could be used to gain the absolute transcript count. By a concentration series, HCR was shown to recover the absolute number of yEVenus transcripts at the working concentration that was the same as that for saturating detection with sFISH ($65\,\mathrm{nM}$). This indicates that the major limitation of the single fluorophore method for achieving high de-

tection efficiency is the percentage of inactive fluorophores in the probe solution. Based other projects in the lab, it was believed that strand displacement could function as a kinetic barrier to FISH probe binding. To explore this hypothesis, we designed a set of HCR probes that were fully matched, but had a mismatch with the target at either the first one or two nucleotides of the stem where displacement would initiate. This experiment revealed that a single mismatch could substantially reduce HCR amplification and a second mismatch completely suppressed amplification for an otherwise fully matched probe. To exploit this specificity, circular RNA became the next target. RNA circles are not commonly found in yeast, however, there are 11 genes with two introns that were demonstrated to express circular transcripts of the middle exon when exposed to heat shock. We demonstrated that the expression of both intron retention isoforms and circular isoforms of SUS1 increases with incubation at $37\,^{\circ}\mathrm{C}$ over $30\,\mathrm{min}$ intervals for two hours. This is the first demonstration of direct single molecule detection of a circular isoform that cannot be uniquely identified by conventional smFISH the sole example of which is CDR1 transcripts[86], the transcript that serves as the miR-7 sponge.

Isoform detection was also extended to an organism with much more complexity in its transcriptome, *C. elegans*. Unlike yeast, very often genes in *C. elegans* often contain many introns and can express alternate splicing variations based on location in the tissue and other factors. Of particular interest to our collaborators was the NURF1 gene, which they expect to have different spatial distributions of NURF1 isoforms based on the location in the germline of the worm. Three sets of HCR probes were chosen to target junctions that allow the identification of five transcript isoforms that are annotated on Wormbase. Furthermore, it was determined that the detection of Alexa488 and Cy3 is much more difficult than Cy5 in *C. elegans*. Therefore to create a more cost effective method an indirect labeling system was chosen so that each binding site contains a toehold for a universal Cy5 probe and a displacer which can deconstruct the amplified polymer. Using strand displacement, the signal can be reset after acquisition and hybridization can be performed sequentially with a

single Cy5 probe. The displacement reaction followed by an incubation in stringent wash buffer can fully suppress the fluorescence of the previous round of hybridization. This represents the first sequential barcoding system to include HCR for repeated hybridization, specificity, and amplification.

## REFERENCES


[1] C. G. Kevil, L. Walsh, F. S. Laroux, T. Kalogeris, M. B. Grisham, and J. Alexander, "An improved, rapid northern protocol," *Biochemical and biophysical research communications*, vol. 238, no. 2, pp. 277–279, 1997.

[2] D. C. Rio, "Reverse transcription–polymerase chain reaction," *Cold Spring Harbor Protocols*, vol. 2014, no. 11, pdb–prot080887, 2014.

[3] S. Bustin, "Quantification of mrna using real-time reverse transcription pcr (rt-pcr): Trends and problems," *Journal of molecular endocrinology*, vol. 29, no. 1, pp. 23–39, 2002.

[4] K. Kurimoto, Y. Yabuta, Y. Ohinata, Y. Ono, K. D. Uno, R. G. Yamada, H. R. Ueda, and M. Saitou, "An improved single-cell cdna amplification method for efficient high-density oligonucleotide microarray analysis," *Nucleic acids research*, vol. 34, no. 5, e42–e42, 2006.

[5] P. A. McGettigan, "Transcriptomics in the rna-seq era," *Current opinion in chemical biology*, vol. 17, no. 1, pp. 4–11, 2013.

[6] M. Li, I. X. Wang, Y. Li, A. Bruzel, A. L. Richards, J. M. Toung, and V. G. Cheung, "Widespread rna and dna sequence differences in the human transcriptome," *science*, p. 1 207 018, 2011.

[7] W. Zheng, L. M. Chung, and H. Zhao, "Bias detection and correction in rna-sequencing data," *BMC bioinformatics*, vol. 12, no. 1, p. 290, 2011.

[8] W. M. Freeman, S. J. Walker, and K. E. Vrana, "Quantitative rt-pcr: Pitfalls and potential," *Biotechniques*, vol. 26, pp. 112–125, 1999.

[9] J Huggett, K Dheda, S Bustin, and A Zumla, "Real-time rt-pcr normalisation; strategies and considerations," *Genes and immunity*, vol. 6, no. 4, pp. 279–284, 2005.

[10] F. Tang, C. Barbacioru, Y. Wang, E. Nordman, C. Lee, N. Xu, X. Wang, J. Bodeau, B. B. Tuch, A. Siddiqui, *et al.*, "Mrna-seq whole-transcriptome analysis of a single cell," *Nature methods*, vol. 6, no. 5, p. 377, 2009.

[11] F. Tang, C. Barbacioru, S. Bao, C. Lee, E. Nordman, X. Wang, K. Lao, and M. A. Surani, "Tracing the derivation of embryonic stem cells from the inner cell mass by single-cell rna-seq analysis," *Cell stem cell*, vol. 6, no. 5, pp. 468–478, 2010.

[12] S. Qiu, S. Luo, O. Evgrafov, R. Li, G. P. Schroth, P. Levitt, J. A. Knowles, and K. Wang, "Single-neuron rna-seq: Technical feasibility and reproducibility," *Frontiers in genetics*, vol. 3, p. 124, 2012.

[13] A. Ståhlberg and M. Bengtsson, "Single-cell gene expression profiling using reverse transcription quantitative real-time pcr," *Methods*, vol. 50, no. 4, pp. 282–288, 2010.

[14] A.-E. Saliba, A. J. Westermann, S. A. Gorski, and J. Vogel, "Single-cell rna-seq: Advances and future challenges," *Nucleic acids research*, gku555, 2014.

[15] S. Islam, U. Kjällquist, A. Moliner, P. Zajac, J.-B. Fan, P. Lönnerberg, and S. Linnarsson, "Characterization of the single-cell transcriptional landscape by highly multiplex rna-seq," *Genome research*, vol. 21, no. 7, pp. 1160–1167, 2011.

[16] Z. Wang, M. Gerstein, and M. Snyder, "Rna-seq: A revolutionary tool for transcriptomics," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 57–63, 2009.

[17] F. Tang, C. Barbacioru, E. Nordman, B. Li, N. Xu, V. I. Bashkirov, K. Lao, and M. A. Surani, "Rna-seq analysis to capture the transcriptome landscape of a single cell," *Nature protocols*, vol. 5, no. 3, pp. 516–535, 2010.

[18] R. H. Singer and D. C. Ward, "Actin gene expression visualized in chicken muscle tissue culture by using in situ hybridization with a biotinated nucleotide analog," *Proceedings of the National Academy of Sciences*, vol. 79, no. 23, pp. 7331–7335, 1982.

[19] M. J. Levesque, P. Ginart, Y. Wei, and A. Raj, "Visualizing snvs to quantify allele-specific expression in single cells," *Nature methods*, vol. 10, no. 9, pp. 865–867, 2013.

[20] R. Jungmann, M. S. Avendaño, J. B. Woehrstein, M. Dai, W. M. Shih, and P. Yin, "Multiplexed 3d cellular super-resolution imaging with dna-paint and exchange-paint," *Nature methods*, vol. 11, no. 3, p. 313, 2014.

[21] M. J. Rust, M. Bates, and X. Zhuang, "Sub-diffraction-limit imaging by stochastic optical reconstruction microscopy (storm)," *Nature methods*, vol. 3, no. 10, pp. 793–796, 2006.

[22] B. J. Beliveau, A. N. Boettiger, M. S. Avendaño, R. Jungmann, R. B. McCole, E. F. Joyce, C. Kim-Kiselak, F. Bantignies, C. Y. Fonseka, J. Erceg, *et al.*, "Single-molecule super-resolution imaging of chromosomes and in situ haplotype visualization using oligopaint fish probes," *Nature communications*, vol. 6, 2015.

[23] J. Lu and A. Tsourkas, "Imaging individual micrornas in single mammalian cells in situ," *Nucleic acids research*, vol. 37, no. 14, e100–e100, 2009.

[24] E. Lubeck and L. Cai, "Single-cell systems biology by super-resolution imaging and combinatorial labeling," *Nature methods*, vol. 9, no. 7, pp. 743–748, 2012.

[25] E. Lubeck, A. F. Coskun, T. Zhiyentayev, M. Ahmad, and L. Cai, "Single-cell in situ rna profiling by sequential hybridization," *Nature methods*, vol. 11, no. 4, pp. 360–361, 2014.

[26] K. H. Chen, A. N. Boettiger, J. R. Moffitt, S. Wang, and X. Zhuang, "Spatially resolved, highly multiplexed rna profiling in single cells," *Science*, vol. 348, no. 6233, aaa6090, 2015.

[27] A. Raj, C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi, "Stochastic mrna synthesis in mammalian cells," *PLoS biology*, vol. 4, no. 10, e309, 2006.

[28] D. Zenklusen, D. R. Larson, and R. H. Singer, "Single-rna counting reveals alternative modes of gene expression in yeast," *Nature structural & molecular biology*, vol. 15, no. 12, p. 1263, 2008.

[29] G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, *et al.*, "Rna velocity of single cells," *Nature*, vol. 560, no. 7719, p. 494, 2018.

[30] A. D. Edelstein, M. A. Tsuchida, N. Amodaj, H. Pinkard, R. D. Vale, and N. Stuurman, "Advanced methods of microscope control using micromanager software," *Journal of biological methods*, vol. 1, no. 2, 2014.

[31] B. E. Saleh, M. C. Teich, and B. E. Saleh, *Fundamentals of photonics*. Wiley New York, 1991, vol. 22.

[32] R. Wagner and W. Tomlinson, "Coupling efficiency of optics in single-mode fiber components," *Applied optics*, vol. 21, no. 15, pp. 2671–2688, 1982.

[33] J. Xu and R. Stroud, *Acousto-optic devices: principles, design, and applications*. Wiley-Interscience, 1992, vol. 12.

[34] S. Bradbury and B. Bracegirdle, "Introduction to light microscopy," in *Introduction to light microscopy*, 1998.

[35] D. Axelrod, T. P. Burghardt, and N. L. Thompson, "Total internal reflection fluorescence," *Annual review of biophysics and bioengineering*, vol. 13, no. 1, pp. 247–268, 1984.

[36] A. Cornea and P. M. Conn, *Fluorescence Microscopy: Super-Resolution and other Novel Techniques*. Elsevier, 2014.

[37] M. Tokunaga, N. Imamoto, and K. Sakata-Sogawa, "Highly inclined thin illumination enables clear single-molecule imaging in cells," *Nature methods*, vol. 5, no. 2, pp. 159–161, 2008.

[38] K. L. Ellefsen, J. L. Dynes, and I. Parker, "Spinning-spot shadowless tirf microscopy," *PLoS One*, vol. 10, no. 8, e0136055, 2015.

[39] W. Zong, X. Huang, C. Zhang, T. Yuan, L.-l. Zhu, M. Fan, and L. Chen, "Shadowless-illuminated variable-angle tirf (siva-tirf) microscopy for the observation of spatial-temporal dynamics in live cells," *Biomedical optics express*, vol. 5, no. 5, pp. 1530–1540, 2014.

[40] A. Negrean and H. D. Mansvelder, "Optimal lens design and use in laser-scanning microscopy," *Biomedical optics express*, vol. 5, no. 5, pp. 1588–1609, 2014.

[41] J. Vangindertael, R Camacho, W Sempels, H Mizuno, P Dedecker, and K. Janssen, "An introduction to optical super-resolution microscopy for the adventurous biologist," *Methods and applications in fluorescence*, vol. 6, no. 2, p. 022 003, 2018.

[42] S Bolte and F. Cordelieres, "A guided tour into subcellular colocalization analysis in light microscopy," *Journal of microscopy*, vol. 224, no. 3, pp. 213–232, 2006.

[43] R. Heintzmann and C. J. Sheppard, "The sampling limit in fluorescence microscopy," *Micron*, vol. 38, no. 2, pp. 145–149, 2007.

[44] R. Roy, S. Hohng, and T. Ha, "A practical guide to single-molecule fret," *Nature methods*, vol. 5, no. 6, p. 507, 2008.

[45] T. Ha, T. Enderle, D. Ogletree, D. S. Chemla, P. R. Selvin, and S. Weiss, "Probing the interaction between two single molecules: Fluorescence resonance energy transfer between a single donor and a single acceptor," *Proceedings of the National Academy of Sciences*, vol. 93, no. 13, pp. 6264–6268, 1996.

[46] G. Weber, "Rotational brownian motion and polarization of the fluorescence of solutions," in *Advances in protein chemistry*, vol. 8, Elsevier, 1953, pp. 415–459.

[47] A. C. Albrecht, "Polarizations and assignments of transitions: The method of photoselection," *Journal of Molecular Spectroscopy*, vol. 6, pp. 84–108, 1961.

[48] D. M. Jameson and W. H. Sawyer, "[12] fluorescence anisotropy applied to biomolecular interactions," in *Methods in enzymology*, vol. 246, Elsevier, 1995, pp. 283–300.

[49] J. Goodman, "Introduction to fourier optics," 2008.

[50] Y. Shechtman, S. J. Sahl, A. S. Backer, and W. Moerner, "Optimal point spread function design for 3d imaging," *Physical review letters*, vol. 113, no. 13, p. 133 902, 2014.

[51] F. Mueller, A. Senecal, K. Tantale, H. Marie-Nelly, N. Ly, O. Collin, E. Basyuk, E. Bertrand, X. Darzacq, and C. Zimmer, "Fish-quant: Automatic counting of transcripts in 3d fish images," *Nature methods*, vol. 10, no. 4, p. 277, 2013.

[52] A. Raj and S. Tyagi, "Detection of individual endogenous rna transcripts in situ using multiple singly labeled probes," in *Methods in enzymology*, vol. 472, Elsevier, 2010, pp. 365–386.

[53] R. E. Thompson, D. R. Larson, and W. W. Webb, "Precise nanometer localization analysis for individual fluorescent probes," *Biophysical journal*, vol. 82, no. 5, pp. 2775–2783, 2002.

[54] D. Perrone and P. Favaro, "A clearer picture of blind deconvolution," *arXiv preprint arXiv:1412.0251*, 2014.

[55] W. H. Richardson, "Bayesian-based iterative method of image restoration," *JOSA*, vol. 62, no. 1, pp. 55–59, 1972.

[56] L. B. Lucy, "An iterative technique for the rectification of observed distributions," *The astronomical journal*, vol. 79, p. 745, 1974.

[57] D.-Y. Tsai, Y. Lee, and E. Matsuyama, "Information entropy measure for evaluation of image quality," *Journal of digital imaging*, vol. 21, no. 3, pp. 338–347, 2008.

[58] K. Smith, Y. Li, F. Piccinini, G. Csucs, C. Balazs, A. Bevilacqua, and P. Horvath, "Cidre: An illumination-correction method for optical microscopy," *Nature methods*, vol. 12, no. 5, p. 404, 2015.

[59] G. M. Wadsworth, R. Y. Parikh, J. S. Choy, and H. D. Kim, "Mrna detection in budding yeast with single fluorophores," *Nucleic acids research*, vol. 45, no. 15, e141–e141, 2017.

[60] R. Y. Parikh and H. D. Kim, "The effect of an intervening promoter nucleosome on gene expression," *PloS one*, vol. 8, no. 5, e63072, 2013.

[61] H. D. Kim and E. K. O'Shea, "A quantitative model of transcription factor–activated gene expression," *Nature structural & molecular biology*, vol. 15, no. 11, pp. 1192–1198, 2008.

[62] A. Kaffman, I. Herskowitz, R. Tjian, and E. K. O'Shea, "Phosphorylation of the transcription factor pho4 by a cyclin-cdk complex, pho80-pho85," *Science*, vol. 263, no. 5150, pp. 1153–1156, 1994.

[63] H. Youk, A. Raj, and A. van Oudenaarden, "Imaging single mrna molecules in yeast," *Methods in enzymology*, vol. 470, pp. 429–446, 2010.

[64] T. Trcek, J. A. Chao, D. R. Larson, H. Y. Park, D. Zenklusen, S. M. Shenoy, and R. H. Singer, "Single-mrna counting using fluorescent in situ hybridization in budding yeast," *nature protocols*, vol. 7, no. 2, pp. 408–419, 2012.

[65] C. E. Aitken, R. A. Marshall, and J. D. Puglisi, "An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments," *Biophysical journal*, vol. 94, no. 5, pp. 1826–1835, 2008.

[66] M. A. Sheff and K. S. Thorn, "Optimized cassettes for fluorescent protein tagging in saccharomyces cerevisiae," *Yeast*, vol. 21, no. 8, pp. 661–670, 2004.

[67] R. N. Spear, S. Li, E. V. Nordheim, and J. H. Andrews, "Quantitative imaging and statistical analysis of fluorescence in situ hybridization (fish) ofaureobasidium pullulans," *Journal of microbiological methods*, vol. 35, no. 2, pp. 101–110, 1999.

[68] A. E. Dodson and J. Rine, "Heritable capture of heterochromatin dynamics in saccharomyces cerevisiae," *Elife*, vol. 4, e05007, 2015.

[69] S. M. Shaffer, M.-T. Wu, M. J. Levesque, and A. Raj, "Turbo fish: A method for rapid single molecule rna fish," *PloS one*, vol. 8, no. 9, e75120, 2013.

[70] V. Pelechano, W. Wei, and L. M. Steinmetz, "Extensive transcriptional heterogeneity revealed by isoform profiling," *Nature*, vol. 497, no. 7447, p. 127, 2013.

[71] C. Jiang and B. F. Pugh, "A compiled and systematic reference map of nucleosome positions across the saccharomyces cerevisiae genome," *Genome Biol*, vol. 10, no. 10, R109, 2009.

[72] T.-L. To and N. Maheshri, "Noise can induce bimodality in positive transcriptional feedback loops without bistability," *Science*, vol. 327, no. 5969, pp. 1142–1145, 2010.

[73] B. Lemieux, N. Laterreur, A. Perederina, J.-F. Noël, M.-L. Dubois, A. S. Krasilnikov, and R. J. Wellinger, "Active yeast telomerase shares subunits with ribonucleoproteins rnase p and rnase mrp," *Cell*, vol. 165, no. 5, pp. 1171–1181, 2016.

[74] M. Zuker, "Mfold web server for nucleic acid folding and hybridization prediction," *Nucleic acids research*, vol. 31, no. 13, pp. 3406–3415, 2003.

[75] A. Edelstein, N. Amodaj, K. Hoover, R. Vale, and N. Stuurman, "Computer control of microscopes using micromanager," *Current protocols in molecular biology*, vol. 92, no. 1, pp. 14–20, 2010.

[76] J. Högbom, "Aperture synthesis with a non-regular distribution of interferometer baselines," *Astronomy and Astrophysics Supplement Series*, vol. 15, p. 417, 1974.

[77] M Oakes, J. Aris, J. Brockenbrough, H Wai, L Vu, and M Nomura, "Mutational analysis of the structure and localization of the nucleolus in the yeast saccharomyces cerevisiae," *The Journal of cell biology*, vol. 143, no. 1, pp. 23–34, 1998.

[78] T. Laroche, S. G. Martin, M. Gotta, H. C. Gorham, F. E. Pryde, E. J. Louis, and S. M. Gasser, "Mutation of yeast ku genes disrupts the subnuclear organization of telomeres," *Current biology*, vol. 8, no. 11, pp. 653–657, 1998.

[79] S Heyden and M Ortiz, "Investigation of the influence of viscoelasticity on oncotripsy," *Computer Methods in Applied Mechanics and Engineering*, vol. 314, pp. 314–322, 2017.

[80] P. Jorgensen, N. P. Edgington, B. L. Schneider, I. Rupeš, M. Tyers, and B. Futcher, "The size of the nucleus increases as yeast cells grow," *Molecular biology of the cell*, vol. 18, no. 9, pp. 3523–3532, 2007.

[81] S. Trumtel, I. Léger-Silvestre, P.-E. Gleizes, F. Teulieres, and N. Gas, "Assembly and functional organization of the nucleolus: Ultrastructural analysis of saccharomyces cerevisiae mutants," *Molecular biology of the cell*, vol. 11, no. 6, pp. 2175–2189, 2000.

[82] T. B. Hansen, J. Kjems, and C. K. Damgaard, "Circular rna and mir-7 in cancer," *Cancer research*, vol. 73, no. 18, pp. 5609–5612, 2013.

[83] S. P. Barrett, P. L. Wang, and J. Salzman, "Circular rna biogenesis can proceed through an exon-containing lariat precursor," *Elife*, vol. 4, 2015.

[84] H. M. Choi, J. Y. Chang, L. A. Trinh, J. E. Padilla, S. E. Fraser, and N. A. Pierce, "Programmable in situ amplification for multiplexed imaging of mrna expression," *Nature biotechnology*, vol. 28, no. 11, p. 1208, 2010.

[85] P. L. Wang, Y. Bao, M.-C. Yee, S. P. Barrett, G. J. Hogan, M. N. Olsen, J. R. Dinneny, P. O. Brown, and J. Salzman, "Circular rna is expressed across the eukaryotic tree of life," *PloS one*, vol. 9, no. 3, e90859, 2014.

[86] X. Zeng, W. Lin, M. Guo, and Q. Zou, "A comprehensive overview and evaluation of circular rna detection tools," *PLoS computational biology*, vol. 13, no. 6, e1005420, 2017.

# VITA

## EDUCATION

Georgia Institute of Technology 2010-2018

Ph.D. in Physics December 2018

Thesis: Uniquely quantifying highly similar RNA transcripts at the single molecule level

M.S. in Physics 2017

B.S. in Physics with Honors 2012

Georgia State University 2007-2008

University of Florida 2000-2001

## AWARDS

Georgia Institute of Technology

Amelio Travel Award, $2,000 2017

Best Graduate Student Poster, SERYM (Southeast Regional Yeast Meeting) 2017

G.A.A.N.N. Fellowship, DOE, Molecular Biophysics Training Program, $19,500 January 2015 January 2016

G.A.A.N.N. Fellowship, DOE, Molecular Biophysics Training Program, $19,500 January 2014 January 2015

Weatherly Travel Award, $500 2014

Graduate Teaching Assistant Award (Finalist for campus-wide TA of the Year) 2014

Georgia State University

Deans Scholarship Key Award 2008

Faculty Scholar Award in 3 semesters January 2007 January 2008

**PEER-REVIEWED PUBLICATIONS**

Garcia, P.D., Wadsworth, G. M., Kim, H. D., and Zakian, V; Untitled study of the mislocalization of TLC1 transcripts (in preparation, submitting 2018)

Gable M. Wadsworth, Harold D. Kim; FISHing on a budget (Submitted, by request of the editor, Methods in Molecular Biology series, Microbial Systems Biology)

Wadsworth, G. M., Parikh, R. Y. and Kim, H. D. (2018). Dual-probe RNA FRET-FISH in Yeast. Bio-protocol 8(11): e2867. DOI: 10.21769/BioProtoc.2867. 2018

Wadsworth, G. M., Parikh, R. Y. and Kim, H. D. (2018). Single-probe RNA FISH in Yeast. Bio-protocol 8(11): e2868. DOI: 10.21769/BioProtoc.2868. 2018

Wadsworth, G. M., Parikh, R. Y., Choy, J. S., and Kim, H. D. (2017). mRNA detection in budding yeast with single fluorophores. Nucleic acids research, 45(15), e141-e141. DOI: 10.1093/nar/gkx568 2017

**ORAL PRESENTATIONS**

Single Fluorophore In Situ Hybridization in Budding Yeast, Physics Forum, Georgia Institute of Technology 2016

Single Fluorophore In Situ Hybridization in Budding Yeast, Molecular Biophysics Seminar, Georgia Institute of Technology 2016

Single-Probe Fluorescence In Situ Hybridization (FISH) in Budding Yeast, Platform: DNA Replication and Transcription, Biophysical Society Meeting 2016

Sensitive Single Fluorophore Detection of mRNA Transcripts in Yeast, GT-Emory Yeast Club meeting, Emory University 2016

Quantification of RNA Transcript level in single yeast cells, Physics of Living Systems Lunch&Learn, Georgia Institute of Technology 2015

Quantification of RNA Transcript level in single yeast cells, Molecular Biophysics Seminar, Georgia Institute of Technology 2015

Quantification of RNA Transcript level in single yeast cells, GT-Emory Yeast Club meeting, Emory University 2015

Quantifying Gene Expression at the mRNA Level, Physics of Living Systems Lunch&Learn, Georgia Institute of Technology 2013

**POSTER PRESENTATIONS**

Highly Specific Circular RNA Quantification In Single Yeast Cells, Biophysical Society Meeting 2018

Highly Specific mRNA Transcript Quantification In Budding Yeast Via Strand Displacement Induced Hybridization Chain Reaction, BPS Thematic Meeting, Single-Cell Biophysics: Measurement, Modulation, and Modeling, Academia Sinica 2017

mRNA Transcript Quantification in Yeast with Single-Molecule FRET, SERYM (Southeast Regional Yeast Meeting), University of Alabama 2017

mRNA Transcript Quantification in Yeast with Single-Molecule FRET, Biophysical Society Meeting 2017

Single Fluorophore In Situ Hybridization in Budding Yeast, iPoLS, Harvard Universiy 2016

Single Fluorophore In Situ Hybridization in Budding Yeast, Soft Matter Lunch&Posters, Georgia Institute of Technology 2016

Single fluorophore quantification of mRNA over ten expression levels, iPoLS, Ludwig Maximilians Universitat 2014

**REVIEWS**

Reviewer, Current Opinion in Biotechnology 2016

**FEATURED**

Prabhune, M. ”To find what’s unique, stop averaging.” (2018): 149-152. 2018